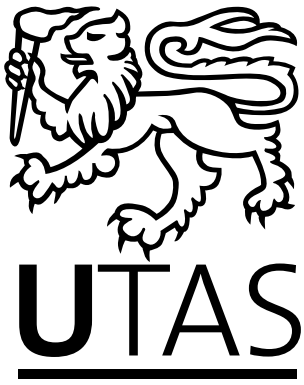
EVALUATING THE ROLE OF MIXED FREQUENCY REAL-TIME WEATHER DATA IN ECONOMIC FORECASTS

by

Thomas Goodwin, BNatResEcon (Hons), GDipEcon

Submitted in partial fulfilment of the requirements
for the Degree of Doctor of Philosophy

Tasmanian School of Business and Economics
University of Tasmania
November, 2017



This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Signed:

Thomas Goodwin

Date: November 5, 2017

ABSTRACT

This thesis examines the role of weather forecasts in economic decision making and proposes a state-space model for evaluating multi-horizon forecasts. Our first application sets the scene by using mixed frequency real-time weather data to improve the predictive performance of retail sales forecasts. Our findings illustrate that significant improvements in predictive accuracy can be achieved when weather variables are incorporated into a mixed frequency forecasting model. However, once we move to a real-time environment, substituting meteorological forecasts for observed weather variables, meteorological forecast uncertainty leads to a somewhat muted improvement in predictive accuracy. The implication is that real-time data constraints are a limitation of forecasting models that use weather variables as an input. This points to a role for forecast evaluation and the need to better understand the structure of forecasting error. Our state-space model of multi-horizon forecasts is based on the forecast revision structure implied by forecast rationality. The parameters of this model allow us to identify rational forecast revision components, that we call information content, as well as residual forecasting error components, that we call implicit forecasting error. A key contribution of this thesis is the demonstration of how our model based approach to forecast evaluation nests existing test based approaches as special cases. We illustrate our proposed multi-horizon forecast evaluation approach using two empirical applications. First, we employ our model based approach to examine the structure of forecasting error in multi-horizon meteorological forecasts of daily temperature. Multi-horizon temperature forecasts are found to contain multiple sources of forecasting error. The structure of meteorological forecasting error depends on the length of the forecast horizon. Second, we employ our model to examine the information content of multi-horizon electricity demand forecasts for the New South Wales region of the Australian National Electricity Market. We discover that the information content of electricity demand forecasts depends not only on the length of the horizon at which the forecasts are made, but also the time of day the forecasts are made.

ACKNOWLEDGEMENTS

A depth of gratitude is owed to the following people who have supported me in one way or another during the course of this endeavour:

Professor Mardi Dungey, for the enthusiasm with which you entertained the many tangential lines of enquiry I proposed through the course of this work. Unvaryingly, such an indulgence was swiftly shelved, allowing me to refocus on more immediate, unresolved questions.

Dr Jing Tian, for confronting my often undisciplined enthusiasm with the rigour and precision necessary to realise our intended objective. Your critical eye will inform my inner voice during future endeavours, and I will be all the better for that.

Sense–T, for the generous scholarship that made this work possible.

Shan, for the sunshine, lollipops, and rainbows you bring to sometimes cold, bland, and colourless days.

Molly, Jethro, and Elliott: may these efforts inspire your direction and trajectory.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

> "... forecasts possess no intrinsic value. They acquire value through their ability to influence the decisions made by users of the forecasts." (Murphy, 1993)

Weather has an important bearing on economic activity. Spurred on by the threat of climate change, a burgeoning literature has appeared concerned with measuring the effect of exogenous variation in weather on economic activity. Dell et al. (2014) provide a comprehensive review of what they term The New Climate-Economy Literature. This new literature uses panel regression models to identify the net effect of weather events on economic activity. It is important to distinguish between climate, which is defined as a distribution of outcomes (e.g. temperature, rainfall), and weather, which is defined as a particular realisation from the climate distribution (Dell et al., 2014; Hsiang, 2016). Weather variation is estimated to influence economic activity across a broad range of sectors including agriculture, industrial production, services, energy, and health.

Quite apart from the issue of climate change, it seems logical to conclude that

decision makers, informed by a well identified understanding of the effect of weather on economic activity, will begin to show an increased level of interest in the use of weather forecasts to predict economic outcomes. As weather forecasts are inherently uncertain, they represent a source of information as well as a source of error for economic decision making. In order to understand the economic value of weather forecasts we need a better understanding the information, as well as the error, contained in weather forecasts.

The goal of this thesis is to provide greater insight into how economic decision makers can best utilise forecasts that are subject to revision. Often, forecasts are revised multiple times before the event being forecast is realised. Ideally this revision process will introduce new information into the forecasts, subsequently reducing the potential for forecasting error. If revised forecasts contain more information, and less error; should decision making be delayed, and if so for how long? Suppose it is Monday and we would like to make a decision that depends on the temperature next Saturday. Meteorological agencies produce forecasts for a range of weather events, at forecast horizons of several hours, days, or even weeks. While it may be desirable to make this decision today, perhaps our decision can be delayed, at a cost, until later in the week. The passage of time will allow meteorologists to gather new information, and revise their expectation of next Saturday's temperature. By waiting until later in the week, we can use revised, and potentially more reliable temperature forecasts. Whether or not to delay our decision involves a trade-off between the cost of delay, and the potential gain from more reliable forecast information. Knowing how reliable forecasts are at each forecast horizon would allow us to make our decision at the optimal horizon. Understanding the structure of the forecast revision process should provide insight into the amount of information

and the amount of error contained in the multi-horizon forecasts.

When data revisions are well behaved they behave like rational forecasting errors (Aruoba, 2008); but what do badly behaved forecast revisions look like? The behaviour of rational forecasting error is well described from a theoretical perspective. Stemming from the rational expectations hypothesis (Muth, 1961), rational forecasting error is equivalent to unpredictable information relevant to the target variable. Let y_t denote the target variable of interest, and let $\hat{y}_{t|t-h}$ denote a forecast of the target produced at horizon h. The following regression is commonly used to test forecast rationality

$$y_t = \alpha_0 + \alpha_1 \hat{y}_{t|t-h} + \epsilon_{t|t-h} \tag{1.1}$$

where $\epsilon_{t|t-h}$ is assumed to possess the usual Gauss Markov properties. A test of the joint null hypothesis that $\alpha_0 = 0$ and $\alpha_1 = 1$ is essentially a test of unbiasedness. This test can be used to asses forecast rationality (Mincer and Zarnowitz, 1969) and the behaviour of data revisions (Mankiw et al., 1984; Mankiw and Shapiro, 1986). Rejection of the null hypothesis provides evidence against rationality, yet non-rejection is not sufficient evidence of rationality.

Taking the second moments of (1.1) we have

$$\text{Var}(y_t) = \text{Var}(\hat{y}_{t|t-h}) + \text{Var}(\epsilon_{t|t-h}) \geq \text{Var}(\hat{y}_{t|t-h}) \tag{1.2}$$

where Var denotes the variance. We expect $\text{Var}(\epsilon_{t|t-h})$ to decline when the forecast horizon becomes shorter as new information about the target will become available to the forecaster. That is, $\text{Var}(\epsilon_{t|t-h})$ will be non-decreasing in h. Following the same logic we expect $\text{Var}(\hat{y}_{t|t-h})$ to be non-increasing in h. These properties can be used to derive rationality tests for multi-horizon forecasts (Jeong and Maddala, 1991; Patton and Timmermann, 2012). Rejection

of the null hypothesis in these tests provide evidence against rationality, yet non-rejection is not sufficient evidence of rationality.

Nordhaus (1987) developed a test of forecast rationality that explicitly considers the forecast revision process. The null hypothesis is that forecast revisions should be unpredictable and the multi-horizon forecast series should follow a random walk. Predictable forecast revisions indicate that the forecaster has not efficiently used available information. Rejection of the null hypothesis for this test provides evidence against rationality, yet non-rejection is not sufficient evidence of rationality; leaving open the question of what badly behaved forecast revisions look like.

The thrust of forecast rationality research has been concerned with testing whether forecasts are rational, or not, and developing ever more sophisticated tests with which to do this (Stekler, 2002). We raise a more open ended, and in many ways a more interesting question: if forecasts are not rational, why not? We are certainly not the first to raise such a question. Lovell (1986) laments that the doctrine of rational expectations does not hold up under empirical scrutiny. Around the same time, and in characteristic style,[1] John Muth wrote an obscure paper proclaiming that rational expectations may not be the best characterisation of the forecasting process. Much like the original Muth (1961) paper, Muth (1985) went largely unnoticed for many years. Davies and Lahiri (1995) considered forecasts that are not entirely rational by modelling the multi-horizon forecasting process. Using a panel approach involving three-dimensional data, Davies and Lahiri (1995, 1999) allow forecasts to contain bias and multiple sources of forecasting error. Their approach is somewhat

[1]For a discussion of persuasive writing and writing style within the economics profession see Sent (2002) and McCloskey (1998).

restricted however, requiring data on multi-horizon forecasts, produced by multiple forecasters, for the same target variable series.

Modelling the way in which initial releases of economic data are revised has been an active area of research for decades (see Croushore (2011); Jacobs and Van Norden (2011) for a review). Data revisions have typically been modelled as either adding news or reducing the amount of noise contained in initial data releases. Motivated by evidence that data revisions are not well behaved, and often exhibit complex structures (Aruoba, 2008; Siklos, 2008), Jacobs and Van Norden (2011) developed a model of data revisions that allows data releases to contain both news and noise components. Motivated by Muth (1985) and Jacobs and Van Norden (2011) we explore whether a model of forecast revisions, that may not be entirely rational, can be used to provide insight into how economic decision makers should best utilise forecasts that are subject to later revision.

Chapter 2 provides a simple illustration of how weather data can be exploited for economic advantage. The influence of temperature on fresh produce sales is used as an example in this chapter. As weather data is available at a higher frequency than the weekly sales data, we use MIxed DAta Sampling (MIDAS) regression models to project weekly sales onto daily temperature variables. Forecasts from MIDAS models are compared with forecasts from a benchmark forecasting model that does not consider temperature, and forecasts from a regression model than only considers average weekly temperature. We achieve encouraging improvements in sales forecast performance when observed daily temperature variables are included in the forecasting model. When observed daily temperature variables are replaced by meteorological forecasts, improvements in forecast performance are less encouraging. This finding illustrates

the role that uncertainty plays when weather forecasts are used as an input to other forecasting models.

In Chapter 3 we develop a model of multi-horizon forecasts based on the forecast revision structure implied by forecast rationality. We then use this model as a novel approach for evaluating the performance of multi-horizon forecasts. Using a modelling approach to evaluate forecast performance has a number of advantages over test-based forecast evaluation approaches. Evaluating forecasts using a dichotomous reject, do not reject, rationality test fails to allow for the multiple, sometimes complex, sources of forecasting error. The parameters of our model allow us to identify rational forecast revision components, that we call information content, as well as residual forecasting error components, that we call implicit forecasting error. As an illustration of our approach we evaluate the performance of multi-horizon maximum daily temperature forecasts for Melbourne, Australia, over the period February 2009 to December 2014.

In Chapter 4 we employ the forecast evaluation model proposed in Chapter 3 to analyse the revision structure of electricity demand forecasts for the New South Wales region of the Australian National Electricity Market (NEM) over the period July 2011 to July 2015. Electricity demand forecasts are produced by the regulator in order to give market participants the forward guidance needed to ensure that adequate supply bids are available to meet dispatch load requirements. Electricity demand forecasts are revised half-hourly in the lead up to dispatch. The structure of this revision process should reveal the amount of information contained in the demand forecasts at each half-hourly pre-dispatch period. Weather is an important driver of electricity demand. The Australian Energy Market Operator (AEMO) use historical and forecast

weather data as inputs to their statistical forecasting models. Weather variables of interest include temperature, humidity, wind speed, and solar radiation (AEMO, 2014). While the pre-dispatch electricity demand forecasts are revised half-hourly, not all of the weather forecasts used as an input to these forecasts are available at the same frequency. Rather than evaluate weather forecasts directly, we evaluate the revisions process of the AEMO electricity demand forecasts. Findings in this chapter suggest that the arrival of information useful for electricity demand forecasting is lumpy, which manifests as distinct peaks in the incremental value of electricity demand forecast revisions.

We conclude this thesis in Chapter 5, where we present a brief summary of our results, along with a discussion of areas for future research.

Chapter 2

Using mixed frequency data to forecast fresh produce demand

2.1 Introduction

The availability of ready to eat fresh produce (washed, cut, and packaged) provides consumers with the convenience of healthy, easy to prepare meals throughout the year. One example of this is bagged refrigerated salad, which is available throughout the year at relatively consistent price and quality. As consumers grow increasingly time poor and health conscious, retailers and their suppliers will be under increased pressure to maintain a continuous supply of conveniently packaged fresh produce. Scheduling a consistent supply of fresh produce requires accurate forecasts of future sales. The accuracy of these forecasts is particularly important as fresh produce is highly perishable.

Sales of bagged salad may be affected by weather conditions. For instance, it is expected that the volume of sales will be high when the temperature is warm and low when the temperature is cool. This seasonal pattern is not due to supply factors as production occurs throughout the year in agro-climatically

diverse regions. Thompson and Wilson (1999) examined the seasonal pattern of bagged salad sales by including temperature variables in a demand model. This temperature augmented model was able to explain seasonal fluctuations in demand, with the seasonal pattern found to be more pronounced at higher latitudes, i.e. those regions that experience a greater variation in temperature between seasons. Despite appreciating the potential influence of daily temperature fluctuations, Thompson and Wilson (1999) calculated monthly average temperatures from daily data so as to match the monthly frequency of the salad sales data. Using monthly average temperatures means that potentially useful information contained daily temperature fluctuations may be masked by the averaging process. For example, warmer than average temperature on weekends may increase salad sales by more than similarly warm weather on weekdays, as salad products complement outdoor dining experiences such as BBQs.

The ability to explain weekly salad sales using daily temperature variables is potentially useful for retailers and their suppliers. Variation in weekly sales around broader seasonal patterns may be related, at least in part, to the variation in actual temperature around seasonal patterns. This is because seasonal climate patterns have been shown influence people's beliefs, while weather influences their actual behaviour (Hsiang, 2016). That is, consumers may plan to eat lighter meals during summer, yet switch to heartier options during weather that is unseasonably cool. Our aim is to determine whether daily temperature data can be used to improve forecasts of weekly salad sales. Specifically, we incorporate daily meteorological forecasts into a mixed-frequency forecasting model of weekly salad sales.

The desire to exploit potential relationships between data sampled at different

frequencies, such as weekly sales and daily temperature, motivated the development of MIxed DAta Sampling (MIDAS) regression models. The MIDAS models introduced by Ghysels et al. (2004) and Ghysels et al. (2007), provide an approach to modelling and forecasting with data sampled at different frequencies by allowing higher frequency predictors to be directly included within a regression model used to forecast a low frequency variable. MIDAS regressions are similar to distributed lag models, yet more flexible, as the higher frequency lags may follow a number of different shapes determined by a data driven function that is characterised by only a small number of parameters. The small number of parameters needed to estimate this function preserve degrees of freedom while allowing for the inclusion of a potentially large number of lags.

MIDAS regressions have been successfully used to improve the predictive performance of quarterly and annual macroeconomic forecasts, such as GDP growth (Clements and Galvão, 2008, 2009), inflation (Monteforte and Moretti, 2013), and private consumption (Dreger and Kholodilin, 2013; Duarte et al., 2017), using monthly, weekly, and even daily data. Recent studies utilising mixed frequency data include forecasts of oil prices (Baumeister et al., 2015), government budgets (Ghysels and Ozkan, 2015), and unemployment (Smith, 2016). To the best of our knowledge this is the first time MIDAS regressions models have been used to forecast retail sales.

We aim to determine whether forecasting models that contain daily temperature information can be used to improve forecasts of weekly salad sales. Our results indicate that daily temperature does a good job of explaining weekly salad sales. Forecasts from models containing temperature as a predictor variable show large reductions in forecasting error relative to a base model. Once we

move to a real-time data environment however, where observed daily temperatures are replaced with meteorological temperature forecasts, the reduction in forecasting error is somewhat muted. This finding highlights the consequences of introducing meteorological forecasting error into the sales forecasting model.

2.2 Data

Our aim is to improve weekly retail salad sales forecasts using meteorological forecasts of maximum daily temperature. The sales data, made available to us by a large Australian retail chain, include the weekly quantity and revenue of bagged salad products sold in over 700 individual stores. There are 105 weeks in the series from 18-December 2011 to 15-December 2013. Individual retailers within the chain carry as many as 60 different types of bagged salad leaf products. Sales quantity is recorded as the number bags sold during the week ending Sunday. We convert the quantity of bags sold into kilograms of salad leaf sold using the individual product specifications. We compute price per kilogram by dividing weekly revenue by the kilograms of salad leaf sold. We focus on weekly salad leaf sales in 142 individual stores located in the state of Victoria, Australia. These stores were chosen as they represent 80 per cent of the total quantity of salad leaf sales in the region covered by our temperature forecast data.

Figure 2.1 shows that sales of bagged salad exhibit a marked seasonal pattern, with a high sales volume when the temperature is warm and low sales volume when the temperature is cool. We augment the weekly salad sales data with higher frequency (daily) temperature observations and meteorological forecasts (degrees Celsius) accessed from Stern (2015). In order to facilitate the real-

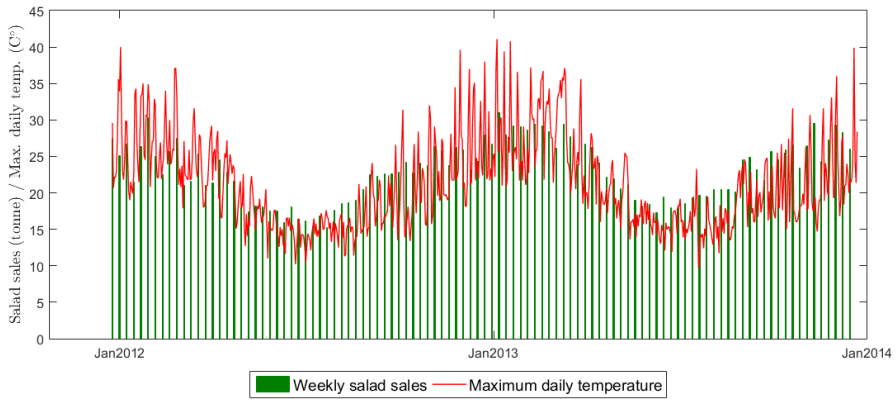


Figure 2.1: Weekly sales of salad leaf and observed maximum daily temperature in Victoria, Australia.

Notes to Figure 2.1: Sales data was made available by an Australian retail chain under the terms of a confidentiality agreement with the University of Tasmania, and includes the aggregate quantity of all salad leaf products sold over the sample period. Temperature data are from the Australian Bureau of Meteorology Melbourne Regional Office weather station (http://www.bom.gov.au/climate/data/).

Table 2.1: RMSE in meteorological maximum daily temperature forecasts, for Melbourne Australia.

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$
RMSE	1.52	1.76	1.96	2.14	2.32	2.65	3.01

Notes to Table 2.1: Values reported in the table are the Root Mean Squared Error (RMSE) over the full sample period. Observed maximum daily temperature data are from the Australian Bureau of Meteorology Melbourne Regional Office weather station (http://www.bom.gov.au/climate/data/). Forecasts of maximum daily temperature were retrieved from http://www.weather-climate.com/.

time forecasting exercise the daily temperature forecast series is constructed by collecting temperature forecasts for Monday through Sunday, as would have been available at the beginning of each sales week. This means that for each low frequency variable (weekly sales) we have seven high frequency variables (daily temperature forecasts). Standing at the beginning of each week the forecast horizon of the meteorological forecasts range from 1-day (where Monday is the target) out to 7-days (where Sunday is the target). Table 2.1 shows that the quality of meteorological forecasts, as measured by root mean square forecasting error (RMSE), declines as the forecast horizon increases. The forecasting error in the meteorological forecasts introduces error into the salad sales forecasting model.

We differentiate between the effect of climate on salad sales and the effect of weather on salad sales. This distinction is important as seasonal climate patterns have been shown to influence consumer beliefs, while weather influences observed consumer behaviour (Hsiang, 2016). We differentiate climate from weather in our data by computing the departure of daily temperature values from the climatological mean for that time of year. The climatological mean for each day of the year is estimated by fitting a Fourier model to observed maximum daily temperature observations for Melbourne, Australia, over the period 1910 to 2010. The climatological mean for each day of the year is assumed to follow the estimated Fourier series. Observed maximum daily temperature observations were obtained from the Australian Bureau of Meteorology. In the sections that follow maximum daily temperature data and meteorological forecasts refer to departures from the climatological mean.

2.3 Mixed frequency forecasting models

Our objective is to produce weekly forecasts of salad sales, denoted S_t, using temperature variables that are sampled at a daily frequency, denoted $W_t^{(d)}$. The subscript t corresponds to the rate at which the low frequency variable is sampled (weekly). The superscript d is the number of times the high frequency variable is sampled between $t-1$ and t (daily). For our example $d = 7$, the number of days in the week.

Before examining the effect of temperature on weekly salad sales we begin with a model that ignores temperature. The performance of this model gives us a benchmark against which to test the forecasting performance of models that include weather variables. The base model of weekly salad sales is adapted

from a dynamic model of fast moving consumer goods in Fok et al. (2007). The forecasting model relates the first difference of log sales to last week's log sales, the first difference of log price, and last week's log price, given by

$$\Delta \ln S_t = \beta_0 + \beta_1 \ln S_{t-1} + \beta_2 \Delta \ln P_t + \beta_3 \ln P_{t-1} + \varepsilon_t, \tag{2.1}$$

where $\ln P_t$ denotes the per kilogram log price of salad and Δ is the first-differencing operator. This model specification was chosen as it captures the dynamics of sales data (Fok et al., 2007).

All models assume that forecasters have knowledge of the salad price during the week being forecast. Price varies from week to week due to product promotions and the discounting of soon to expire product. Retailers will have accurate knowledge of price related promotions for the week being forecast, however price fluctuations related to other forms of discounting could potentially cause endogeneity issues in the forecasting model. We ignore potential endogeneity issues in what follows.

In the subsections that follow we consider three different approaches for handling data sampled at mixed frequencies within a regression framework: using a simple average to temporally aggregate the higher frequency data so that it matches the lower frequency variable; estimating weights for the high frequency variables within a MIDAS regression; and allowing the coefficients to be unrestricted in a U-MIDAS model.

2.3.1 Flat weight regression

One way to approach the mixed frequency data problem is to compute average weekly temperature from the daily data so that the variables on the left and

right hand side of the forecasting model have the same frequency. When data have the same frequency we can estimate the model using ordinary least squares. Using the simple weekly average of the daily temperature data, a forecasting model for weekly salad sales can be written as

$$\Delta \ln S_t = \beta_0 + \beta_1 \ln S_{t-1} + \beta_2 \Delta \ln P_t + \beta_3 \ln P_{t-1} + \beta_4 \sum_{i=0}^{6} \frac{1}{7} W_{t-i/7}^{d} + \varepsilon_t. \quad (2.2)$$

The problem with this approach is that each of the daily temperature variables receive the same weight. As all of the high frequency variables enter the forecasting model with the same weight this approach is often referred to as a flat-weight regression. Using a flat-weight regression ignores potentially useful information in the high frequency data, and may result in estimation bias as temperature is likely to be more important for explaining sales on certain days of the week. This suggests that a regression approach with flexible weights may be a more appropriate forecasting model.

2.3.2 Exponential Almon MIDAS

MIDAS regressions allows us to directly incorporate the high frequency temperature variables into the forecasting model. The MIDAS model for combining daily temperature data with weekly sales data, for 1-week ahead forecasting, is

$$\Delta \ln S_t = \beta_0 + \beta_1 \ln S_{t-1} + \beta_2 \Delta \ln P_t + \beta_3 \ln P_{t-1} + \beta_4 B(L^{1/d}; \theta) W_t^{(d)} + \varepsilon_t. \quad (2.3)$$

The MIDAS lag polynomial $B(L^{1/d}; \theta)$, projects weekly sales S_t onto a series of daily temperature variables W_t^d, using the Almon lag weight function

$$B(L^{1/d}; \theta) = \sum_{j=0}^{6} b(j; \theta) L^{1/d}, \quad (2.4)$$

where the lag operator is defined as

$$L^{1/d}(W_t^d) = W_{t-j/d}^d, \tag{2.5}$$

and the function $b(j;\theta)$ is the weight given to the W_t^d daily temperature variables. The weight function is parameterised using an exponential Almon lag polynomial specification

$$b(j;\theta) = \frac{exp(\theta_1(j+1) + \theta_2(j+1)^2)}{\sum_{j=0}^{6} exp(\theta_1(j+1) + \theta_2(j+1)^2)} \tag{2.6}$$

that permits the weights to take on a range different shapes depending in the values $\theta \equiv \{\theta_1, \theta_2\}$.

The projection of weekly salad sales S_t onto the series of daily temperature forecasts can be seen more clearly when (2.3) is expressed as

$$\begin{aligned}\Delta \ln S_t &= \beta_0 + \beta_1 \ln S_{t-1} + \beta_2 \Delta \ln P_t + \beta_3 \ln P_{t-1} \\ &+ \beta_4[b(1;\theta)W_t^{(d)} + b(2;\theta)W_{t-1/d}^{(d)} + \ldots + b(6;\theta)W_{t-5/d}^{(d)} + b(7;\theta)W_{t-6/d}^{(d)}] + \varepsilon_t\end{aligned} \tag{2.7}$$

where $W_t^{(d)}, W_{t-1/d}^{(d)}, \ldots, W_{t-5/d}^{(d)}, W_{t-6/d}^{(d)}$ refer to daily temperature variables for Sunday, Saturday, ..., Tuesday, Monday, during the week in which the salad sales S_t occur.

The restrictions imposed by the exponential Almon polynomial weights are useful when there is a large number of high frequency regressors. For instance, were we to combine hourly temperature data and weekly sales data the higher frequency data would be sampled $d = 168$ times between $t-1$ and t. As we consider only $d = 7$ the exponential Almon polynomial weights may be unnecessarily restrictive.

2.3.3 Unrestricted MIDAS

An alternative modelling approach is to place no restrictions on the high frequency predictors and estimate a separate coefficient for each W_t^d. An unrestricted or U-MIDAS forecasting model for weekly salad sales can be written as

$$\Delta \ln S_t = \beta_0 + \beta_1 \ln S_{t-1} + \beta_2 \Delta \ln P_t + \beta_3 \ln P_{t-1} + \sum_{i=0}^{6} \beta_i W_{t-i/d}^d + \varepsilon_t \qquad (2.8)$$

where each $W_{t-i/d}^d$ variable has a separate slope coefficient and the model can be estimated using ordinary least squares.

2.4 Empirical results

Forecasts are produced using each of the models described in Section 2.3. Our first set of weather augmented forecasts are produced using observed maximum daily temperature values both fit to the model, and as predictor variables. This provides a benchmark with which to determine the influence of any error contained in the meteorological forecasts. The next set of weather augmented forecasts are subject to real-time data constraints. Instead of using observed temperature as a predictor, pseudo out of sample forecasts are produced using meteorological forecasts that would have been available at the start of the week as predictor variables.

The week ending 25-December 2011 marks the start of the estimation period. The week ending 30-June 2013 marks the end of the initial estimation period. The estimation period is update recursively on a weekly basis until the week ending 8-December 2013. This represents 24 out of sample forecasts, covering

the second half of the 2013 calendar year, using a sample of weekly salad sales data covering 104 weeks.

The salad sales forecasts are evaluated in levels with respect to the actual weekly sales quantity in the dataset. We examine two different measures of forecast performance. We evaluate the relative performance by comparing RMSE for the alternative models with RMSE of the base model. This allows us to determine the influence of weather variables on forecast performance, and to determine the influence of using different weights for the higher frequency variables. We also report mean absolute percentage error (MAPE) values to provide an indication of the mean deviation from the quantity of stock that would be required to meet weekly salad sales.

Due to parameter estimation uncertainty the Diebold and Mariano (1995) test is inappropriate for evaluating between MIDAS regression models (Baumeister et al., 2015). Ghysels and Ozkan (2015) propose using the Giacomini and White (2006) test and we follow this approach.

2.4.1 Parameter estimates

In Table 2.2 we show parameter estimates for the base model, flat weight model, exponential Almon model, and the U-MIDAS model. Parameters are estimated on the full 104 weeks of aggregated salad sales. Sales figures are based on the equilibrium quantity of products sold. The base model parameter estimates are all statistically significantly different from zero. Parameter estimates for the price variables have the expected sign. For example, we expect that the parameter estimate for the difference of log price should have a negative value. Including temperature variables in the base model substan-

tially improves model fit, according to reported R^2 values. In the flat-weight regression model a 1 C° increase in average weekly temperature is estimated to result in a 0.026 percentage point increase weekly sales. Turning to the MIDAS regressions, parameter estimates for $\{\theta_1, \theta_2\}$ are not shown as they determine the shape of the exponential Almon weights yet do not have an economic interpretation. Using the exponential Almon regression model a 1C° increase in the weighted average of weekly temperature is estimated to result in a 0.02 percentage point increase weekly sales, less than the estimate for the flat weight regression. Only the parameter estimates for Tuesday and Wednesday temperature are statistically significantly different from zero in the U-MIDAS regression model. The importance of temperature during the beginning of the week for explaining weekly salad sales is also evident in the estimated exponential Almon weights. The estimated weights for Tuesday and Wednesday are the highest in the exponential Almon model. Figure 2.2 compares the estimated weights for the exponential Almon and coefficient estimates for the U-MIDAS model.

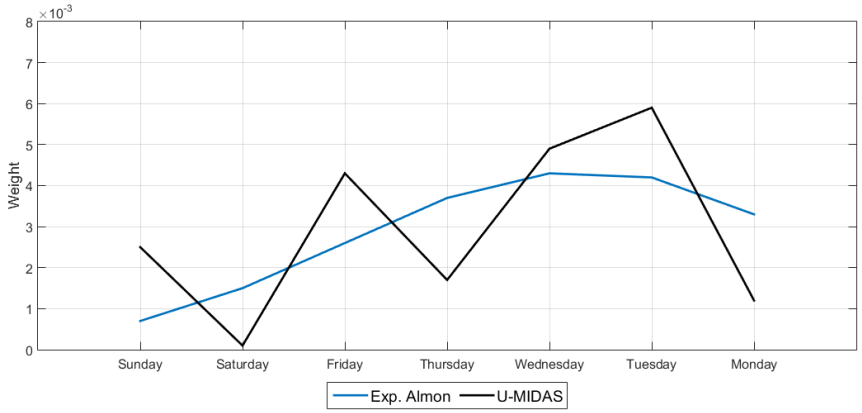


Figure 2.2: Estimated weights for maximum daily temperature lags in the MIDAS regressions.

2.4.2 Predictive performance

Table 2.3 reports on the performance of forecasts for the aggregate quantity of weekly salad sales across all 183 stores. This scenario may be important for a retail chain that uses a centralised distribution centre for bulk ordering and storage before dispatching to individual stores. We first examine the forecast performance of alternative models using observed maximum daily temperature values. This assumes that the forecaster has access to error free meteorological forecasts. The left panel of Table 2.3 shows that all of the models that include observed temperature as predictors have lower RMSEs than forecasts produced by the base model. The U-MIDAS model achieves the highest RMSE reduction among the weather augmented models. The exponential Almon regression models show a similar reduction in RMSE relative to the base model, and they out perform the flat weight model.

The right panel of Table 2.3 shows forecasting models that incorporate meteorological forecasts do not reduce RMSE by as much as forecasting models that use observed maximum daily temperature. Using this real-time data constraint demonstrates the implications of introducing the forecasting error contained in meteorological forecasts into the salad sales forecasts. Nevertheless, U-MIDAS models still perform the best when forecasts of maximum temperature are used to predict weekly salad sales.

We conduct tests for the significance of the forecast improvements shown in Table 2.3 using the approach proposed by Giacomini and White (2006). The columns labelled GW p-value in Table 2.3 report the p-value of these tests where the weather augmented models are compared to the base model that does not include temperature. When forecasting aggregate sales for all stores

the only forecasting model found to have significant predictive gains over the base model is the U-MIDAS model that utilises meteorological forecasts. One explanation for this finding could be that consumers base their purchases on expected temperature.

Table 2.2: Parameter estimates for the base model and mixed-frequency regression models for all stores.

Parameters	Basline model		Flat weights		Exp. Almon		U-MIDAS	
Coefficients	Estimate	Std. error	Estimate	Std. error	Estimate	Std. error	Estimate	Std. error
Intercept	4.715	(1.797)	5.321	(1.371)	7.538	(1.661)	7.390	(1.756)
MIDAS slope			0.026	(0.003)	0.020	(0.003)		
Sun. temp.							0.002	(0.002)
Sat. temp.							0.000	(0.002)
Fri. temp.							0.004	(0.002)
Thu. temp.							0.002	(0.002)
Wed. temp.							0.005	(0.002)
Tue. temp.							0.006	(0.003)
Mon. temp.							0.001	(0.002)
lag log sales	-0.173	(0.055)	-0.257	(0.043)	-0.257	(0.045)	-0.248	(0.048)
diff. log price	-1.007	(0.501)	-1.177	(0.382)	-1.126	(0.415)	-1.088	(0.438)
lag log price	-0.991	(0.513)	-0.920	(0.391)	-1.662	(0.454)	-1.643	(0.479)
R^2	0.087		0.471		0.513		0.538	

Notes to Table 2.2: Standard errors are in parentheses. Sales data was made available by an Australian retail chain under the terms of a confidentiality agreement with the University of Tasmania, and includes the aggregate quantity of all salad leaf products sold over the sample period. Temperature data are from the Australian Bureau of Meteorology Melbourne Regional Office weather station (http://www.bom.gov.au/climate/data/). Models of salad sales that include temperature as an explanatory variable have a higher goodness of fit.

Table 2.3: Predictive performance of forecasts for aggregate sales.

	Observed temperature			Forecast temperature		
	RMSE ratio	MAPE	GW p-value	RMSE ratio	MAPE	GW p-value
Flat weights	**0.79**	5.03%	0.43	**0.88**	5.23%	0.47
Exp. Almon	**0.67**	4.47%	0.09	1.05	7.14%	0.35
U-MIDAS	**0.66**	4.36%	0.06	**0.73**	4.77%	0.02

Notes to Table 2.3: Sales data was made available by an Australian retail chain under the terms of a confidentiality agreement with the University of Tasmania, and includes the aggregate quantity of all salad leaf products sold over the sample period. Temperature data are from the Australian Bureau of Meteorology Melbourne Regional Office weather station (http://www.bom.gov.au/climate/data/). RMSE ratios less than 1 (bold font) indicate superior predictive performance relative to the baseline model. GW p-value less than 0.05 indicates statistically significant predictive performance relative to the baseline model using the Giacomini and White (2006) test.

Table 2.4: Summary of predictive performance of forecasts for individual store sales.

	Observed temperature						
	Average		Minimum		Maximum		GW p-value
	RMSE ratio	MAPE	RMSE ratio	MAPE	RMSE ratio	MAPE	Freq.
Flat weight	0.91	7.95%	0.69	4.23%	1.20	19.43%	2.11%
Exp. Almon	0.85	7.42%	0.62	4.08%	1.13	19.87%	24.65%
U-MIDAS	0.87	7.60%	0.61	4.67%	1.26	22.19%	16.20%

	Forecast temperature						
	Average		Minimum		Maximum		GW p-value
	RMSE ratio	MAPE	RMSE ratio	MAPE	RMSE ratio	MAPE	Freq.
Flat weight	0.96	8.14%	0.76	4.38%	1.16	19.79%	1.41%
Exp. Almon	1.03	9.02%	0.47	4.64%	1.57	20.18%	5.63%
U-MIDAS	0.90	7.72%	0.67	4.49%	1.24	21.59%	13.38%

Notes to Table 2.4: Sales data was made available by an Australian retail chain under the terms of a confidentiality agreement with the University of Tasmania, and includes the aggregate quantity of all salad leaf products sold over the sample period. Temperature data are from the Australian Bureau of Meteorology Melbourne Regional Office weather station (http://www.bom.gov.au/climate/data/). RMSE ratios less than 1 (bold font) indicate superior predictive performance (averaged across stores, maximum, and minimum) relative to the baseline model (averaged over all stores). Mean Absolute Percentage Error (MAPE) provides an indication of the absolute deviation from demand that would result if the forecasts were used for inventory control (averaged across stores, maximum, and minimum) GW p-value less than 0.05 indicates statistically significant predictive performance relative to the baseline model using the Giacomini and White (2006) test. The figure reported under the heading GW p-value is the frequency of stores with significant predictive performance relative to the baseline model (averaged across stores, maximum, and minimum).

Table 2.4 reports summary measures of the performance of forecasts for the quantity of weekly salad sales in the 183 individual stores. The top panel shows the performance of forecasts that use observed maximum daily temperature. The bottom panel shows the performance of forecasts that use meteorological forecasts of maximum daily temperature. As with the aggregate forecasts, using the unrestricted MIDAS model to incorporate daily maximum temperature for predicting weekly salad sales outperform flat weights and the exponential Almon weights. The average of the RMSE reductions for individual stores is lower when meteorological temperature forecasts are used as predictors. The range of RMSE reductions for individual stores presented in Table 2.4 and the histograms of RMSE ratios in Figure 2.3 show that the MIDAS model with unrestricted weights performs the most favourably under the real-time data constraint. Comparing the bottom three histograms of Figure 2.3 illustrates that the distribution of individual store RMSE values for the U-MIDAS model sits further below the RMSE ratio of 1 than both the RMSE distributions for the flat weight regression and exponential Almon models.

The MAPE values reported in Table 2.3 and Table 2.4 suggest a similar forecast performance ranking as the RMSE ratio measures. By way of interpretation, consider the MAPE values in Table 2.4 for an unrestricted MIDAS model and individual stores under the real-time data constraint. The MAPE values for this model range from a low of 4.49% to a high of 21.59%. Meanwhile the MAPE value obtained for sales aggregated across all stores is 4.77%. This result suggests there is a potential gain from having a centralised distribution centre handle bulk orders that can be dispatched to individual stores several times per week.

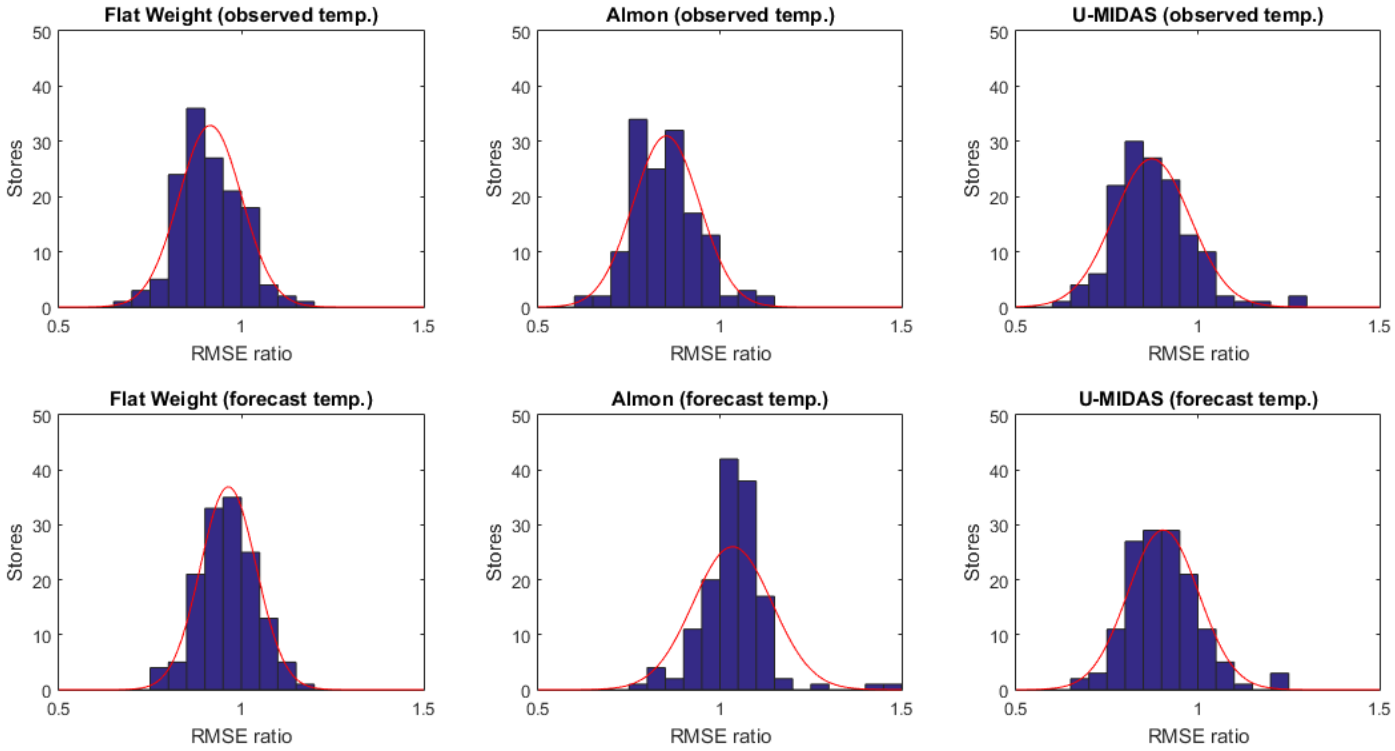


Figure 2.3: Forecasting models that include observed temperature as a predictor have greater predictive performance than models that use temperature forecasts.

Notes to Figure 2.3: Sales data was made available by an Australian retail chain under the terms of a confidentiality agreement with the University of Tasmania, and includes the aggregate quantity of all salad leaf products sold over the sample period. Temperature data are from the Australian Bureau of Meteorology Melbourne Regional Office weather station (http://www.bom.gov.au/climate/data/). Frequency distributions are for stores representing 80 per cent of the total quantity of salad leaf sales in the region covered by the temperature forecast data. RMSE ratios less than 1 indicate superior predictive performance relative to the baseline model.

2.5 Conclusion

This chapter has explored a number of approaches for forecasting weekly salad sales using daily temperature as a predictor. We consider three different approaches for handling data sampled at mixed frequencies within a regression framework: using a simple average to temporally aggregate the higher frequency data so that it matches the lower frequency variable; estimating weights for the high frequency variables within a MIDAS regression; and allowing the coefficients to be unrestricted in a U-MIDAS model. MIDAS regressions have been successfully used to improve the predictive performance of a range of macroeconomic forecasts, commodity price forecasts, and government budget forecasts, among other applications. To the best of our knowledge this is the first time MIDAS regressions models have been used to forecast retail sales.

Although none of the forecasting models we examine show statistically significant improvements in predictive ability, several conclusions can be drawn from our results. Firstly, weather appears to be important for explaining weekly fluctuations in salad sales. Second, including observed temperature variables in the forecasting model leads to greater forecast improvements than using meteorological temperature forecasts, although this difference in performance was not found to be statistically significant. It appears that the forecasting error in meteorological forecasts has implications for the accuracy of forecasting models that include temperature as a predictor.

The consequences of forecasting error for real-time decision making demands consideration of the structure of forecasting error and an understanding of how the structure of forecasting error evolves over the forecast horizon. In Chapter 3 we propose a state space approach to multi-horizon forecast evaluation.

Chapter 3

A state-space approach to multi-horizon forecast evaluation

3.1 Introduction

Forecast evaluation is an important source of information for business and economic decision making, not to mention for forecasters themselves. In this chapter we develop and implement a state space approach for evaluating the revision structure of multi-horizon forecasts. This approach is attractive because it allows us to use a number of properties implied by forecast rationality to decompose multi-horizon forecasts into horizon specific measures of bias, information content, and inefficiency.

It is typical to evaluate the rationality of forecasts. Using the Rational Expectations hypothesis proposed by Muth (1961) as a framework with which to evaluate forecasts sets up a range of testable hypothesis. The first of these is that forecasts should be unbiased. Mincer and Zarnowitz (1969) first proposed a regression approach for testing forecast biasedness. Despite the fact that Holden and Peel (1990) have shown the null hypothesis of this test to be suf-

ficient, yet not necessary, for forecast unbiasedness the Mincer and Zarnowitz (1969) approach is still commonly used in the literature. This approach is referred to as rolling event forecast evaluation as a series of target variables are used to test the biasedness of forecasts produced at a specific horizon. Separate tests are used to evaluate forecasts produced at different horizons.

Rational forecasts are efficient. Forecasts are said to be efficient when forecasters incorporate all available information into their forecasts. Therefore it should not be possible to use information available at the time the forecasts were made to make ex-post forecast improvements. Nordhaus (1987) introduces the concept of weak forecast efficiency as an alternative forecast evaluation approach. Rather than testing whether all available information has been used in the forecasting process, weak form efficiency requires that forecasts for the same target variable have revisions that are independent of past revisions and past forecasting errors. This approach is referred to as fixed event forecast evaluation, as a fixed target variable is used to test the efficiency of a series of multi-horizon forecasts. Clements (1997) proposes a more robust test of weak form efficiency by pooling the series of multi-horizon forecast across multiple target variables. Clements and Taylor (2001) extend this approach to allow for the possibility that forecasting error is not normally distributed. Davies and Lahiri (1995, 1999) use a panel data approach, where multi-horizon forecasts are produced by multiple forecasters, to develop tests of forecast biasedness and forecast efficiency.

As well as implying unbiasedness and efficiency, rationality has a number of implications for the second moment bounds of forecast variables. Jeong and Maddala (1991) use the implication that forecast variables should have a smaller variance than the target variable as a test of forecast rationality.

Isiklar and Lahiri (2007) compare the variance of forecast revisions between two horizons with the change in the variance of forecasting error between the same two horizons. The rationale for comparing these second moments is that efficient forecast revisions should lead to a reduction in forecasting error. Forecast inefficiency is suspected when the variance of forecast revisions are found to be larger than the change in the variance of forecasting error between the same horizons. Patton and Timmermann (2012) propose an additional suite of inequality tests based on properties implied by the rationality of multi-horizon forecasts. They propose ten variance and covariance bounds tests that may be used to identify different forms of departure from the rationality implications.

Our model based approach to forecast evaluation has a number of advantages over existing tests of forecast evaluation. We decompose multi-horizon forecast revisions into a measure of information content and a measure of forecast inefficiency. The second moment bounds implied by forecast rationality are used to model the rational component of multi-horizon forecasts and we allow for additional sources of forecasting error by introducing separate horizon specific error components. Model selection criteria are used to determine which model specification best characterises the multi-horizon forecast series under evaluation. Horizon specific parameter estimates provide a measure of forecast information content and forecast inefficiency, which is a richer source of information than simply rejecting, or failing to reject, a hypothesis of forecast rationality.

We apply our model based forecast evaluation approach to a real-time dataset of temperature forecasts for Melbourne, Australia. Our results suggest that the weather forecasts contain bias and multiple sources of forecasting error. Additionally, our results suggest that the contribution of multi-horizon forecasting

error components may vary both along the forecast horizon, and between different sub-samples of the series being forecast.

The remainder of this chapter is structured as follows. In Section 3.2 we propose a model of multi-horizon forecasts that contain multiple sources of forecasting error. In Section 3.3 we cast our model of multi-horizon forecasts in state space form. In Section 3.4 we compare our forecast evaluation approach with existing forecast evaluation approaches. We evaluate multi-horizon weather forecasts as an empirical illustration of our approach in Section 3.5. Section 3.6 concludes.

3.2 Multifarious error in multi-horizon forecasts

In this section we develop a model of multi-horizon forecasts that contain multiple sources of error. We begin by outlining our assumptions and describing our most general model specification. In the subsections that follow we explain how the dynamics of two different types of forecasting error can be modelled as unobserved components. The two types of forecasting error we consider are rational forecasting error and implicit forecasting error[1]

3.2.1 A model of multi-horizon forecasts

Suppose we are interested in evaluating multi-horizon forecasts of a stochastic univariate process $y \equiv \{y_t; t = 1, 2, \ldots\}$, where y_t belongs to a general class

[1]We make a conscious effort to distance our discussion of rational forecasting error and implicit forecasting error from the news and noise terminology discussed in the data revisions literature, see (Croushore, 2011).

$$\begin{bmatrix} \hat{y}_{t-1|t-h-1} & \hat{y}_{t-1|t-h} & \hat{y}_{t-1|t-h+1} & \cdots & y_{t-1} & y_{t-1} \\ \text{NA} & \hat{y}_{t|t-h} & \hat{y}_{t|t-h+1} & \cdots & \hat{y}_{t|t-1} & y_t \\ \vdots & \vdots & \hat{y}_{t+1|t-h+1} & \cdots & \hat{y}_{t+1|t-1} & \hat{y}_{t+1|t} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ \text{NA} & \text{NA} & \text{NA} & \text{NA} & \cdots & \hat{y}_{t+h|t} \end{bmatrix}$$

Figure 3.1: Revision trapezoid for multi-horizon forecasts.

Notes to Figure 3.1: Later time periods correspond with a movement down the columns and shorter forecast horizons correspond to a movement from left to right across the rows. A movement across the rows also corresponds with later vintages. A vintage contains the most up to date information that would have been available to someone at that time.

of stochastic processes not requiring assumptions of serial independence or covariance stationarity. Let the h horizon expectation $\mathrm{E}(\cdot)$ of y_t have the subscript $t-h$, where $h \geq 0$, to show that it is conditioned on the information set $\mathcal{F}_{t-h}$. This information set is the σ-field $\mathcal{F}_{t-h} \equiv \sigma\{\mathbf{x_{t-h-k}}\ , k \geq 0\}$, where $\mathbf{x_{t-h}}$ is a vector of variables relevant to outcome of y_t. The elements of $\mathbf{x_{t-h}}$ need not be restricted to past and current values of y_t.

At each forecast horizon, $h = 1, 2, \ldots$ forecasters form an expectation about the target variable y_t and use this expectation to produce a forecast $\hat{y}_{t|t-h}$. As the target date approaches the forecaster produces a sequence of forecast revisions, $d_{t|h-1,h} = \hat{y}_{t|t-h-1} - \hat{y}_{t|t-h}$, by updating previous horizon forecasts. Figure 3.1 uses the revision trapezoid format, commonly used in the real-time data literature (Croushore, 2011), to illustrate the structure multi-horizon forecast data we wish to evaluate. We move to later time periods as we move down the columns and we move to shorter forecast horizons as we move left to right across the rows. For instance, in the column on the far right hand side of the table the time period runs from $t-1$ to $t+h$. In the second top row of the table the conditioning on each of the variables runs from time period $t-h$ to time period t. Another way to describe a movement across the rows is to say that we move to later vintages. A vintage contains the most up to date information that would have been available to someone at that time. The

column for the time t vintage on the far right hand side of the table contains the realised target values y_{t-1} and y_t, as well as the forecast values $\hat{y}_{t+h|t}$ for $1, 2, \ldots, h$.

Our aim is to characterise a series of multi-horizon forecasts with a model that contains one or more sources of forecasting error. Our most general model specification decomposes forecasts into the following unobserved components: a dynamic representation of the target variable, $\tilde{y}_t$; rational forecasting error, $\nu_{t|t-h}$; and implicit forecasting error, $\zeta_{t|t-h}$. Bringing all these terms together, we express multi-horizon forecasts, with multiple sources of error, as follows

$$\hat{y}_{t|t-h} = \tilde{y}_t + \nu_{t|t-h} + \zeta_{t|t-h}. \tag{3.1}$$

where we assume that the forecasts are unbiased, such that the unconditional expectation of the unobserved rational and implicit forecasting errors are zero, that is $\mathrm{E}(\nu_{t|t-h}) = \mathrm{E}(\zeta_{t|t-h}) = 0$.

In the subsections that follow we discuss the dynamics of rational multi-horizon forecasting error and implicit forecasting error.

3.2.2 Rational multi-horizon forecasts

Our most restrictive model of multi-horizon forecasts assumes that forecasts are rationally revised at all forecast horizons. The dynamics of rational forecasts and their target variable evolve according to the rational expectations hypothesis of Muth (1961). According to Muth (1961) the only source of error in a rational forecast is unpredictable information, relevant to the target variable, that is realised after the forecast is made. Let ω_i, for $i = h-1, \ldots 0$ denote information relevant to the target variable y_t, realised after the horizon

h forecast $\hat{y}_{t|t-h}$ is made. We are interested in modelling the accumulation of information between the longest forecast horizon and the time at which the target variable is realised. Consider the following target variable model

$$y_t = \tilde{y}_t \tag{3.2}$$

where y_t is the observed value of the target variable and $\tilde{y}_t$ is assumed to be the true value of the target variable at time t. The dynamics of $\tilde{y}_t$ evolve according to

$$\tilde{y}_t = \tilde{y}_{t-1} + \varepsilon_t \tag{3.3}$$

where

$$\varepsilon_t = \xi_t + \sum_{i=0}^{H-1} \omega_i = \sigma_\xi \eta_{\xi,t} + \sum_{i=0}^{H-1} \sigma_{\omega_i} \eta_{\omega_i,t}. \tag{3.4}$$

such that $\eta_{\xi_t} \sim i.i.d.\ \mathrm{N}(0,1)$ and $\eta_{\omega_{i,t}} \sim i.i.d.\ \mathrm{N}(0,1)$. Information relevant to the target variable that was available prior to the commencement of the longest forecast horizon being evaluated is defined as ξ_t. We use H to denote the longest forecast horizon being evaluated. We use h to denote individual forecast horizons, all of which have H as a reference point. Rational multi-horizon forecasts are modelled as

$$\hat{y}^*_{t|t-h} = \tilde{y}_t + \nu_{t|t-h} \tag{3.5}$$

where

$$\nu_{t|t-h} = -\sum_{i=0}^{h-1} \omega_i = -\sum_{i=0}^{h-1} \sigma_{\omega_i} \eta_{\omega_i,t} \tag{3.6}$$

is information relevant to the target variable that was not available when the horizon h forecast was produced. Subtracting a horizon h forecast (3.5) from the observed target (3.2), as follows

$$e^*_{t|t-h} \equiv y_t - \hat{y}^*_{t|t-h} = \sum_{i=0}^{h-1} \omega_i = \sum_{i=0}^{h-1} \sigma_{\omega_i} \eta_{\omega_i,t} \tag{3.7}$$

reveals forecasting error equal to the sum of unpredictable information, relevant to the target variable, realised after the horizon h forecast is made. Rational forecasting error $e^*_{t|t-h} \equiv -\nu_{t|t-h}$ possesses a number of desirable properties: 1) $e^*_{t|t-h}$ is correlated with the target variable as $\text{Cov}(\tilde{y}_t, \nu_{t|t-h}) \neq 0$; 2) $e^*_{t|t-h}$ is uncorrelated with the forecast variable as $\text{Cov}(\hat{y}_{t|t-h}, \nu_{t|t-h}) = 0$; and 3) the variance of rational forecasting error, computed as $\text{E}(\nu^2_{t|t-h}) = \sum_{i=0}^{h-1} \sigma^2_{\omega_i}$, is non-decreasing in h. This last property is intuitively appealing as we expect there to be less relevant information available to forecasters at longer horizons. Less relevant information will cause an increase in forecast uncertainty and an associated increase in the variability of rational forecasting error. Conversely, as the forecast horizon becomes shorter rational forecasts accumulate information which means that the variance of rational forecasts $\text{E}(\hat{y}^2_{t|t-h})$ will be non-decreasing as h decreases.

3.2.3 Implicit forecasting error

Our most general model of multi-horizon forecasts in (3.1) assumes that forecasts contain rational forecasting error as well as horizon specific mistakes at one or more forecast horizons. We refer to these mistakes as implicit forecasting error, as the dynamics of these errors are inspired by the theory of implicit expectations (Mills, 1957). Implicit expectations says that forecasting error is uncorrelated with the target variable $\text{cov}(\tilde{y}_t, \zeta_{t|t-h}) = 0$. We model implicit forecasting error as a horizon specific innovation

$$\zeta_{t|t-h} = \sigma_{\zeta_h} \eta_{\zeta_h,t} \tag{3.8}$$

where $\eta_{\zeta_h,t} \sim i.i.d.\ \text{N}(0,1)$. While still unpredictable, implicit forecasting error is introduced by the forecaster. This implies that implicit forecasting errors are

correlated with their forecasts, that is $\text{Cov}(\hat{y}_{t|t-h}, \zeta_{t|t-h}) \neq 0$. There are a number of ways that forecasters may introduce implicit error into their forecasts. Information used by forecasters may contain measurement error. Additionally, forecasters may use information about the target variable in a way that is inefficient. The inefficient use of information may be unintentional, such as when a forecaster under or over react to the arrival of new information (Isiklar and Lahiri, 2007). At other times a forecaster may intentionally use information in a way that is inefficient in order to smooth forecast revisions (Nordhaus, 1987). For example, a forecaster wishing to cultivate a reputation for producing stable forecasts may be reluctant to make large revisions to previous forecasts, even when presented with new information. We place no restrictions on the pattern of implicit forecasting error over the forecast horizon. Therefore the variance of implicit forecasting error, computed as $\text{E}(\zeta^2_{t|t-h}) = \sigma^2_{\zeta_h}$, may increase or decrease as the forecast horizon shortens.

3.3 A state space representation of multi-horizon forecasts

We model the dynamics of multi-horizon forecasts, containing multiple sources of error, by casting the problem in state space form. The time-invariant state space model consists of a set of measurement equations and a set of transition equations (Durbin and Koopman, 2012)

$$\boldsymbol{y}_t = \boldsymbol{Z}\boldsymbol{\alpha}_t \tag{3.9}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{R}\boldsymbol{\eta}_t. \tag{3.10}$$

The measurement vector, $\boldsymbol{y}_t = [\hat{y}_{t|t-h}, \hat{y}_{t|t-(h-1)}, \ldots, y_t]$, stacks observed multi-

horizon forecast variables on the observed target variable.[2] In our most general model the state vector is partitioned as follows

$$\boldsymbol{\alpha}_t = \begin{bmatrix} \tilde{y}_t & \boldsymbol{\nu}'_{t|t-h} & \boldsymbol{\zeta}'_{t|t-h} \end{bmatrix}' \tag{3.11}$$

which has a length $(1 + h + h)$. The associated measurement equation is

$$\boldsymbol{y}_t = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{Z}_2 & \boldsymbol{Z}_3 \end{bmatrix} \cdot \begin{bmatrix} \tilde{y}_t \\ \boldsymbol{\nu}'_{t|t-h} \\ \boldsymbol{\zeta}'_{t|t-h} \end{bmatrix} \tag{3.12}$$

where $\boldsymbol{Z} = [\boldsymbol{Z}_1 \quad \boldsymbol{Z}_2 \quad \boldsymbol{Z}_3]$ is a partitioned matrix conforming with the unobserved components of the state vector: $\boldsymbol{Z}_1 = \mathbf{1}_{(h+1)}$ (a $(h+1)\times 1$ vector of ones) relates to the target variable component; $\boldsymbol{Z}_2 = \left[\begin{smallmatrix} \boldsymbol{I}_h \\ \mathbf{0} \end{smallmatrix}\right]$ (a $h \times h$ identity matrix, atop a conformably defined vector of zeros) relates to the rational component; and $\boldsymbol{Z}_3 = \left[\begin{smallmatrix} \boldsymbol{I}_h \\ \mathbf{0} \end{smallmatrix}\right]$ (a $h \times h$ identity matrix, atop a conformably defined vector of zeros) relates to the implicit component.[3]

Transition equations describe the dynamics of the unobserved components in terms of the state vector

$$\begin{aligned} \begin{bmatrix} \tilde{y}_t \\ \boldsymbol{\nu}'_{t|t-h} \\ \boldsymbol{\zeta}'_{t|t-h} \end{bmatrix} &= \begin{bmatrix} T_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{T}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_3 \end{bmatrix} \cdot \begin{bmatrix} \tilde{y}_{t-1} \\ \boldsymbol{\nu}'_{t-1|t-1-h} \\ \boldsymbol{\zeta}'_{t-1|t-1-h} \end{bmatrix} \\ &+ \begin{bmatrix} R_1 & \boldsymbol{R}_2 & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{U} \cdot \operatorname{diag}(\boldsymbol{R}_2) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \operatorname{diag}(\boldsymbol{R}_3) \end{bmatrix} \cdot \begin{bmatrix} \eta_{\xi,t} \\ \boldsymbol{\eta}'_{\omega_h,t} \\ \boldsymbol{\eta}'_{\zeta_h,t} \end{bmatrix} \end{aligned} \tag{3.13}$$

The matrices T_1, R_1 and $\boldsymbol{R}_2$ describe the dynamics of the target variable component. The target variable is represented as a Random Walk Model plus ω_i innovations, for $i = H - 1, \ldots 0$

$$\tilde{y}_t = T_1 \cdot \tilde{y}_{t-1} + R_1 \cdot \eta_{\xi,t} + \boldsymbol{R}_2 \cdot \boldsymbol{\eta}'_{\omega,t} \tag{3.14}$$

[2]Where h is the longest horizon at which we observe the multi-horizon forecasts.

[3]It should be possible to deal explicitly with target variables that contain error by adding an extra parameter to the implicit component. So as not to complicate matters we assume that the observed value of the target variable is measured without error. Any measurement error specific to the target variable will be contained in the parameter σ_{ω_0}. This assumption does not impact any of the results that follow.

where $T_1 = 1$, $R_1 = \sigma_\xi$, and $\boldsymbol{R}_2 = [\sigma_{\omega_{h-1}}, \dots \sigma_{\omega_0}]$ are standard deviations of the innovations ω_i.

The matrix $\boldsymbol{T}_2$ is a null matrix[4] and the dynamics of the rational component are completely described by $-\boldsymbol{U} \cdot \text{diag}(\boldsymbol{R}_2)$ in the $\boldsymbol{R}$ matrix. To impose the property that rational forecasting error is the negative sum of ω_i innovations we first specify a $h \times h$ matrix with the elements of the vector $\boldsymbol{R}_2$ on its main diagonal diag($\boldsymbol{R}_2$). Then diag($\boldsymbol{R}_2$) is pre-multiplied by $-\boldsymbol{U}$, a $h \times h$ matrix with zeros below the main diagonal and each of the remaining elements equal to one. This allows us to model the rational component as follows

$$\boldsymbol{\nu}_{t|t-h} = -\boldsymbol{U} \cdot \text{diag}(\boldsymbol{R}_2) \cdot \boldsymbol{\eta}'_{\nu,t} = - \begin{bmatrix} \sigma_{\omega_{h-1}} & \sigma_{\omega_{h-2}} & \dots & \sigma_{\omega_0} \\ 0 & \sigma_{\omega_{h-2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_{\omega_0} \\ 0 & \dots & 0 & \sigma_{\omega_0} \end{bmatrix} \cdot \begin{bmatrix} \eta_{\omega_{h-1}t} \\ \eta_{\omega_{h-2}t} \\ \vdots \\ \eta_{\omega_0 t} \end{bmatrix}. \quad (3.15)$$

Jacobs and Van Norden (2011) use this approach to model consecutive vintages of data revisions as a set of rational forecasting errors.

The matrix $\boldsymbol{T}_3$ is a null matrix and the dynamics of the implicit component are completely described by diag($\boldsymbol{R}_3$) in the $\boldsymbol{R}$ matrix. Implicit forecasting errors are horizon specific innovations, uncorrelated with the target variable, and uncorrelated with each other, so we model the dynamics of the implicit component as follows

$$\boldsymbol{\zeta_{t|t-h}} = \text{diag}(\boldsymbol{R}_3) \cdot \boldsymbol{\eta}'_{\zeta,t} = \begin{bmatrix} \sigma_{\zeta_h} & 0 & \dots & 0 \\ 0 & \sigma_{\zeta_{h-1}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{\zeta_1} \end{bmatrix} \cdot \begin{bmatrix} \eta_{\zeta_h t} \\ \eta_{\zeta_{h-1} t} \\ \vdots \\ \eta_{\zeta_1 t} \end{bmatrix} \quad (3.16)$$

[4]Forecasting error that has a moving average process of at most $MA(h-1)$ may still be consistent with forecast optimality. This is because forecasts from horizons $h-1, \dots 1$ are produced before forecasting error from horizon h is realised (Clements, 1997). We may capture this correlation structure by estimating parameters ρ on matrix $\boldsymbol{T}_2$ and matrix $\boldsymbol{T}_3$. However, our focus is on evaluating forecasting error components that deviate from rationality, so we ignore this type of correlation in the present chapter.

where the elements $\boldsymbol{R}_3 = [\sigma_{\zeta_h}, \sigma_{\zeta_{h-1}}, \ldots \sigma_{\zeta_1}]$ are standard deviations of the implicit innovations.

3.3.1 Estimating the state space model

The rational and implicit components described above can be used as the building blocks for two different models of multi-horizon forecasts:

rational-implicit forecasts represented by

$$\boldsymbol{y}_t = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{Z}_2 & \boldsymbol{Z}_3 \end{bmatrix} \cdot \begin{bmatrix} \tilde{y}_t \\ \boldsymbol{\nu}'_{t|t-h} \\ \boldsymbol{\zeta}'_{t|t-h} \end{bmatrix} \tag{3.17}$$

$$\begin{bmatrix} \tilde{y}_t \\ \boldsymbol{\nu}'_{t|t-h} \\ \boldsymbol{\zeta}'_{t|t-h} \end{bmatrix} = \begin{bmatrix} T_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{T}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_3 \end{bmatrix} \cdot \begin{bmatrix} \tilde{y}_{t-1} \\ \boldsymbol{\nu}'_{t-1|t-1-h} \\ \boldsymbol{\zeta}'_{t-1|t-1-h} \end{bmatrix} + \begin{bmatrix} R_1 & \boldsymbol{R}_2 & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{U} \cdot \mathrm{diag}(\boldsymbol{R}_2) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathrm{diag}(\boldsymbol{R}_3) \end{bmatrix} \cdot \begin{bmatrix} \eta_{\xi,t} \\ \boldsymbol{\eta}'_{\omega,t} \\ \boldsymbol{\eta}'_{\zeta,t} \end{bmatrix}; \tag{3.18}$$

and

rational forecasts represented by

$$\boldsymbol{y}_t = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{Z}_2 \end{bmatrix} \cdot \begin{bmatrix} \tilde{y}_t \\ \boldsymbol{\nu}'_{t|t-h} \end{bmatrix} \tag{3.19}$$

$$\begin{bmatrix} \tilde{y}_t \\ \boldsymbol{\nu}'_{t|t-h} \end{bmatrix} = \begin{bmatrix} T_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{T}_2 \end{bmatrix} \cdot \begin{bmatrix} \tilde{y}_{t-1} \\ \boldsymbol{\nu}'_{t-1|t-1-h} \end{bmatrix} + \begin{bmatrix} R_1 & \boldsymbol{R}_2 \\ \mathbf{0} & -\boldsymbol{U} \cdot \mathrm{diag}(\boldsymbol{R}_2) \end{bmatrix} \cdot \begin{bmatrix} \eta_{\xi,t} \\ \boldsymbol{\eta}'_{\omega,t} \end{bmatrix}. \tag{3.20}$$

Once the multi-horizon forecasts are modelled in state space form we obtain parameter estimates of unobserved multi-horizon forecasting error components with maximum likelihood methods using the Kalman filter.

3.4 Evaluating multi-horizon forecasts

In this section we explain how our model of multi-horizon forecasts can be used as a unified framework for identifying a number of different sources of forecast sub-optimality. Our approach to forecast evaluation involves three steps: first we fit both the rational model and the rational-implicit model to the multi-horizon forecast series under investigation; next we use model selection criteria to choose the model that best characterises the data; and then, we compare parameter estimates from the chosen model with a set of well established properties implied by forecast rationality.

In the subsections that follow we discuss how other forecast evaluation approaches are nested as special cases of our model. In subsection 3.4.1 we discuss internal consistency, and in subsection 3.4.2 we discuss forecast efficiency. As with existing tests of internal consistency and forecast efficiency, our approach assumes that the forecasts being evaluated are unbiased. In subsection 3.4.3 we review the limitations of existing tests of forecast biasedness and propose an alternative state space approach to detect and extract multi-horizon forecasts bias.

3.4.1 Internal consistency

In this subsection we use variance and covariance bound properties implied by forecast rationality as a lens through which to interpret the parameters of our rational and rational-implicit models of multi-horizon forecasts.

Using equations (3.2), (3.5), and (3.7) we can express the variance of rational multi-horizon forecasts as

$$\text{Var}(y_t) = \text{Var}(\hat{y}^*_{t|t-h}) + \text{Var}(e^*_{t|t-h}) \tag{3.21}$$

where the condition

$$\text{Cov}(\hat{y}^*_{t|t-h}, e^*_{t|t-h}) = 0 \tag{3.22}$$

is implied by forecast rationality. As the variance of the target variable $\text{Var}(y_t)$ does not change with changes in the forecast horizon h, a suite of multi-horizon variance and covariance bound properties can be derived from (3.21) and (3.22). Variance and covariance bound properties have been used elsewhere as hypotheses with which to test forecast rationality (Jeong and Maddala, 1991; Patton and Timmermann, 2012). Forecast evaluation approaches that exploit the properties of multi-horizon forecasts as h changes are known as fixed event forecast evaluation approaches.

Patton and Timmermann (2012) propose a suite of multi-horizon forecast evaluation tests based on variance and covariance bound properties implied by forecast rationality. Specifically, they test the monotonicity properties of the second moment bounds of multi-horizon forecast variables. Two sets of forecast evaluation tests are developed. The first set of inequality tests, based on Gourieroux et al. (1982) and Wolak (1987, 1989), assume that y_t is covariance stationary. While the second set of inequality tests, based on Wooldridge and

White (1988) and White (2001) do not require the assumption that y_t is covariance stationarity. The latter tests are more general as they use a central limit theorem for heterogenous, serially dependent processes. However, unit root processes are excluded. Our model based approach explicitly allows y_t to belong to a general class of stochastic processes, including unit root processes, and we draw a link between our approach and the Patton and Timmermann (2012) test by comparing estimates of the parameters σ_{ω_h} and σ_{ζ_h}.

Table 3.1 shows conditions under which the Patton and Timmermann (2012) tests will be satisfied in the model specified in the previous sections of this chapter. The first column of Table 3.1 contains the variance and covariance bound properties that are the basis of rationality tests considered by Patton and Timmermann (2012). The remaining columns in Table 3.1 either contain "$\surd$" or a set of conditions. Looking down the columns under each model specification: cells marked with "$\surd$" indicate that the variance or covariance bound property in the first column is unconditionally satisfied; the remaining cells describe the conditions under which the variance and covariance bounds properties in the first column will be satisfied. The conditions refer to the pattern of σ_ω and σ_{ζ_h} parameters at different forecast horizons. When these conditions are satisfied the null of the relevant Patton and Timmermann (2012) test in the first column will not be rejected (proof of this is presented in Appendix A).

As expected, each row of the column labelled "Rational" in Table 3.1 contains a "$\surd$" indicating that our model of rational multi-horizon forecasts unconditionally satisfies all ten of the variance and covariance bound properties. Therefore, a multi-horizon forecast series that can be characterised by our rational model specification will not reject any of the variance and covariance

bounds tests proposed by Patton and Timmermann (2012).

The patterns we expect the variance of rational forecasts $\text{Var}(\hat{y}^*_{t|t-h})$ and the variance of rational forecasting error $\text{Var}(e^*_{t|t-h})$ to display across forecast horizons (Row 1 and Row 2 of Table 3.1) can be explained in terms of information content in the multi-horizon forecasts. As the forecast horizon h becomes shorter, we expect forecasters to learn more about the target variable, and incorporate more information into their forecasts. Consequently the variance of rational forecasts $\text{Var}(\hat{y}^*_{t|t-h})$ will be non-decreasing as h becomes shorter. The parameters of our model of rational multi-horizon forecasts unconditionally satisfy the variance bound property that $\text{Var}(\hat{y}^*_{t|t-h})$ is non-decreasing as h becomes shorter. Using equations (3.3) and (3.5), we have

$$\text{Var}(\hat{y}^*_{t|t-h}) = \text{Var}(\tilde{y}_{t-1}) + \sigma^2_\xi + \sum_{i=h-1}^{H-1} \sigma^2_{\omega_i}. \tag{3.23}$$

Since the first two terms remain the same as we vary the forecast horizon h, $\text{Var}(\hat{y}^*_{t|t-h})$ is non-decreasing as h becomes shorter.

As previously discussed, we interpret the parameter σ_{ω_h} as information relevant to the target variable. According to the rational expectations hypothesis the only source of error in a multi-horizon forecast $\hat{y}^*_{t|t-h}$ is information relevant to the target variable that will become available after forecasts are made at horizon h. Therefore, at shorter horizons, information accumulation is increasing. Our model of rational multi-horizon forecasts unconditionally satisfies the variance bound property that $\text{Var}(e^*_{t|t-h})$ is non-decreasing in h, given that the variance of (3.7) is

$$\text{Var}(e^*_{t|t-h}) = \sum_{i=0}^{h-1} \sigma^2_{\omega_i}. \tag{3.24}$$

When newly available information is incorporated into a multi-horizon forecast

we say that the forecast is revised. Forecast revisions that contain a larger amount of information will have a larger variance. Under rational expectations it is assumed that all of the information in forecast revisions will be relevant to the target variable. Suppose the revision between a rational forecast made at horizon h and a rational forecast at the longer horizon $h+k$ is $d^*_{t|h,h+k} = \hat{y}^*_{t|t-h} - \hat{y}^*_{t|t-(h+k)}$. Using equation (3.5) we derive the variance of the rational forecast revision as

$$\text{Var}(d^*_{t|h,h+k}) \equiv \sum_{i=h}^{h+k-1} \sigma^2_{\omega_i}. \tag{3.25}$$

Therefore, $\text{Var}(d^*_{t|h,h+k})$ is non-decreasing as the number of horizons over which forecasts are revised increases. While still useful from a testing perspective, the interpretation of the other variance and covariance patterns listed in Table 3.1 are intuitively less obvious.

Parameter estimates from our rational-implicit model of multi-horizon forecasts may lead to variance and covariance patterns that are inconsistent with the properties listed in Table 3.1. The parameters of our rational-implicit model suggest why Patton and Timmermann (2012) tests reject (or fail to reject) the null hypothesis of forecast rationality. Patton and Timmermann (2012) observe, using simulations and an empirical example, that some of their tests reject the null hypothesis of forecast rationality, while others simultaneously fail to reject. The size and power of the inequality tests may be one reason for this mixed testing outcome. We argue that another reason for this outcome may be the presence of multiple sources of forecasting error. The column labelled "Rational-Implicit" lists the conditions under which parameter estimates from the rational-implicit model will be consistent with the variance and covariance bound patterns listed in the first column of Table 3.1.

Table 3.1: Summary of the monotonicity patterns of variance and covariance bounds for different models of multi-horizon forecasts.

	Patton and Timmermann (2012) tests	**Rational**	**Rational-Implicit**
PT1	$\text{Var}(\hat{y}_{t\|t-s}) \geq \text{Var}(\hat{y}_{t\|t-l})$	$\surd$	*iff* $\sum_{i=s}^{l-1} \sigma^2_{\omega_i} \geq \sigma^2_{\zeta_l} - \sigma^2_{\zeta_s}$, $\forall\ l{>}s$
PT2	$\text{Var}(e_{t\|t-s}) \leq \text{Var}(e_{t\|t-l})$	$\surd$	*iff* $\sum_{i=s}^{l-1} \sigma^2_{\omega_i} \geq \sigma^2_{\zeta_s} - \sigma^2_{\zeta_l}$, $\forall\ l{>}s$
PT3	$\text{Var}(d_{t\|s,m}) \leq \text{Var}(d_{t\|s,l})$	$\surd$	*iff* $\sum_{i=m}^{l-1} \sigma^2_{\omega_i} \geq \sigma^2_{\zeta_m} - \sigma^2_{\zeta_l}$, $\forall\ l{>}m$
PT4	$\text{Cov}(\hat{y}_{t\|t-s}, y_t) \geq \text{Cov}(\hat{y}_{t\|t-l}, y_t)$	$\surd$	$\surd$
PT5	$\text{Cov}(\hat{y}_{t\|t-m}, \hat{y}_{t\|t-s}) \geq \text{Cov}(\hat{y}_{t\|t-l}, \hat{y}_{t\|t-s})$	$\surd$	$\surd$
PT6	$\text{Var}(d_{t\|s,l}) \leq 2\text{Cov}(y_t, d_{t\|s,l})$	$\surd$	*iff* $\sum_{i=s}^{l-1} \sigma^2_{\omega_i} \geq \sigma^2_{\zeta_s} + \sigma^2_{\zeta_l}$, $\forall\ l{>}s$ *where* s *is fixed.*
PT7	$\text{Var}(d_{t\|m,l}) \leq 2\text{Cov}(\hat{y}_{t\|t-s}, d_{t\|m,l})$	$\surd$	*iff* $\sum_{i=m}^{l-1} \sigma^2_{\omega_i} \geq \sigma^2_{\zeta_m} + \sigma^2_{\zeta_l}$, $\forall\ l{>}m$ *where* m *is fixed.*
PT8	$\text{Cov}(e_{t\|t-s}, y_t) \leq \text{Cov}(e_{t\|t-l}, y_t)$	$\surd$	$\surd$
PT9	$\text{Cov}(d_{t\|s,m}, \hat{y}_{t\|t-s}) \leq \text{Cov}(d_{t\|s,l}, \hat{y}_{t\|t-s})$	$\surd$	$\surd$
PT10	$\text{Cov}(e_{t\|t-m}, d_{t\|s,m}) \leq \text{Cov}(e_{t\|t-l}, d_{t\|s,l})$	$\surd$	*iff* $\sum_{i=m}^{l-1} \sigma^2_{\omega_i} \geq \sigma^2_{\zeta_m} - \sigma^2_{\zeta_l}$, $\forall\ l{>}m$

Notes to Table 3.1 Var denotes variance and Cov denotes covariance. The generic $t-h$ forecast horizons considered above are $l > m > s$. "$\surd$" indicates that the Patton and Timmermann (2012) variance and covariance bounds properties are unconditionally satisfied. When a model specification does not unconditionally satisfy the properties in the first column we provide a description of the conditions under which the variance and covariance bounds properties will be satisfied.

The "$\surd$" in the rows labelled PT4, PT5, PT8, and PT9 in the column labelled "Rational-Implicit" indicate that forecasts that can be characterised by the rational-implicit model will unconditionally satisfy the covariance bound tests in Column 1. This suggests that forecasts containing implicit forecasting error will fail to reject the null hypothesis of forecast rationality for these Patton and Timmermann (2012) tests, regardless of the pattern exhibited by the variance of the implicit component.

The remaining rows of the column labelled "Rational-Implicit" suggest a set of conditions that the parameters σ_{ω_h} and σ_{ω_h} must possess in order to satisfy the patterns in Column 1 of Table 3.1. Taken together the conditions for tests PT1, PT2, PT3, and PT10 indicate that when the change in the variance of the implicit component along the forecast horizon outweighs the change in the variance of the rational component over the same forecast horizons, forecasts that can be characterised by the rational-implicit model will not satisfy the relevant variance and covariance bound tests. This suggests a number of conditions under which forecasts containing implicit forecasting error may, or may not, fail to reject the null hypothesis of Patton and Timmermann (2012) forecast rationality tests.

Test PT6 and PT7 in Table 3.1 describe a more restrictive set of conditions that depend on the relative size of the variance of the rational and implicit components. The intuition behind these properties is that an increase in the variance of a rational-implicit forecast revision is related to the amount of information relevant to the target variable and the amount of random error. The variance of a rational forecast revision is only related to the amount of information relevant to the target variable. Therefore, the variance of a rational-implicit forecast revision $\text{Var}(d_{t|s,l})$ will be higher than the variance

of a rational forecast revision $\text{Var}(d^*_{t|s,l})$ with the same relevant information. However, the covariance between this rational-implicit forecast revision and the target variable $\text{Cov}(y_t, d_{t|s,l})$ will be the same as the covariance between the rational forecast revision and the target variable $\text{Cov}(y_t, d^*_{t|s,l})$. So the property that — twice the covariance between the forecast revision and the target variable should be at least as large as the variance of the forecast revision — will depend on the implicit component of the forecast revision.

The conditions in tests PT6 and PT7 are more restrictive than the other conditions in Table 3.1, as they indicate that the rational component of a forecast revision between two horizons should be at least as large the sum of the implicit components over the same two horizons. The other inequality conditions only require the variance of the rational component of a forecast revision to be at least as large the difference between the variance of the implicit components over the same two horizons. This difference depends on the direction of the variance or covariance monotonic pattern.

Multi-horizon forecasts that can be characterised by our rational-implicit model may satisfy some of the ten monotonicity properties in Table 3.1. All of these properties are implied by forecast rationality. This means that a rejection of the null hypothesis indicates that the forecasts are not rational, whereas a failure to reject the null is not sufficient evidence of forecast rationality. It is entirely possible for forecasts containing implicit forecasting error to exhibit the well behaved monotonic variance and covariance patterns across multiple forecast horizons that are implied by the rational expectations hypothesis. The practical implication of this result is that the Patton and Timmermann (2012) tests are akin to a test of the internal consistency multi-horizon forecasts implied by forecast rationality (West, 2012). Internal consistency is a

weaker form of forecast rationality than forecast efficiency (Lahiri, 2012). In the next subsection we discuss a number of measures of forecast efficiency and the related concept of forecast information content.

3.4.2 Forecast efficiency

In this subsection we explore the connection between our model based approach to forecast evaluation and the concept of forecast efficiency proposed by Nordhaus (1987), as well as related measures of forecast information content proposed by Isiklar and Lahiri (2007).

The concept of forecast efficiency introduced by Nordhaus (1987) measures the extent to which information is incorporated into forecasts. The two propositions used to test forecast efficiency are: 1) forecasting error at horizon h is independent of all forecast revisions up to horizon $h+1$; and 2) the forecast revision between horizon $h+1$ and horizon h is independent of all forecast revisions up to horizon $h+1$.

Using our model of rational-implicit multi-horizon forecasting error $e_{t|t-h}$ and forecast revisions $d_{t|h+1,h+k}$, we have

$$\text{Cov}(e_{t|t-h}, d_{t|h+1,h+k}) = 0 \tag{3.26a}$$

$$\Rightarrow \text{Cov}\left[(\sum_{i=0}^{h-1} \sigma_{\omega_i}\eta_{\omega_i,t} + \sigma_{\zeta_h}\eta_{\zeta_h,t}), (\sum_{i=h+1}^{h+k-1} \sigma_{\omega_i}\eta_{\omega_i,t} + \sigma_{\zeta_{h+1}}\eta_{\zeta_{h+1},t} - \sigma_{\zeta_{h+k}}\eta_{\zeta_{h+k},t})\right] = 0 \tag{3.26b}$$

implying that forecasting error at horizon h is independent of all forecast revisions up to horizon $(h+1)$ for all $k > 1$. Therefore, multi-horizon forecasts that can be characterised by our rational-implicit model will not reject this first test.

We now check if the rational-implicit multi-horizon forecasts satisfy the second Nordhaus (1987) property of forecast efficiency. Given that

$$\mathrm{Cov}(d_{t|h,h+1}, d_{t|h+1,h+k}) \tag{3.27a}$$

$$= \mathrm{Cov}\left[(\sum_{i=h+1}^{h+k-1} \sigma_{\omega_i}\eta_{\omega_i,t} + \sigma_{\zeta_{h+1}}\eta_{\zeta_{h+1},t} - \sigma_{\zeta_{h+k}}\eta_{\zeta_{h+k},t}), (\sigma_{\omega_h}\eta_{\omega_h,t} + \sigma_{\zeta_h}\eta_{\zeta_h,t} - \sigma_{\zeta_{h+1}}\eta_{\zeta_{h+1},t})\right] \tag{3.27b}$$

$$= \sigma^2_{\zeta_{h+1}}, \tag{3.27c}$$

the forecast revision between horizon $h+1$ and horizon h is not independent of all forecast revisions up to horizon $h+1$ for all $k > 1$. Therefore, multi-horizon forecasts that can be characterised by our rational-implicit model will reject this second test. This finding illustrates that multi-horizon forecasts containing implicit forecasting error are not consistent with forecast efficiency as defined by Nordhaus (1987). The tests proposed by Nordhaus (1987) do not indicate the horizon at which the forecasts are inefficient. The related measure of forecast information content suggested by Isiklar and Lahiri (2007) allow us to determine horizon specific measures of forecast inefficiency.

Isiklar and Lahiri (2007) suggest that when forecasts are optimal, the change in the mean square error between forecasts from horizon h and $h-1$, $\Delta_h \mathrm{MSE}_t = \mathrm{MSE}_{t|t-(h-1)} - \mathrm{MSE}_{t|t-h}$, should be equivalent to the mean square revision of forecasts between horizon h to $h-1$, $\mathrm{MSR}_{t|h-1,h}$. The intuition comes back to the idea that the only source of error in rational forecasts is information relevant to the target variable that will be realised at future h horizons. Forecasting error will decline when a forecaster realises information relevant to the target variable at horizon $h-1$, and efficiently uses this information to revise the forecast they produced at horizon h. This same forecast revision will

increase the amount of information contained in the horizon $h-1$ forecast, relative to the amount of information in the horizon h forecast. This idea is similar to the Patton and Timmermann (2012) variance bound property $\text{Var}(d_{t|h-1,h}) \leq 2\text{Cov}(y_t, d_{t|h-1,h})$, yet the condition proposed by Isiklar and Lahiri (2007) is more restrictive. Using the parameters of our rational-implicit model, a change in the mean square error between the forecast at horizon h and the forecast at horizon $h-1$ can be expressed as follows

$$\Delta_h \text{MSE}_t = \text{E}[e^2_{t|t-h}] - \text{E}[e^2_{t|t-(h-1)}] \tag{3.28a}$$

$$\equiv \sum_{i=0}^{h-1} \sigma^2_{\omega_i} + \sigma^2_{\zeta_h} - \sum_{i=0}^{h-2} \sigma^2_{\omega_i} - \sigma^2_{\zeta_{h-1}} \tag{3.28b}$$

$$= \sigma^2_{\omega_{h-1}} + \sigma^2_{\zeta_h} - \sigma^2_{\zeta_{h-1}} \tag{3.28c}$$

where the first two terms on the right hand side of (3.28c) are the maximum possible forecast improvement between horizons h and $h-1$. The term $\sigma^2_{\omega_{h-1}}$ represents forecast improvement due to the arrival of information relevant to the target variable at horizon $h-1$. The term $\sigma^2_{\zeta_h}$ is implicit forecasting error, specific to the horizon h forecast. Implicit forecasting error $\sigma^2_{\zeta_h}$ disappears as we move to horizon $h-1$. However, the third term, $\sigma^2_{\zeta_{h-1}}$, represents implicit forecasting error contained in the horizon $h-1$ forecast. We model implicit forecasting error as a horizon specific mistake, yet from a practical standpoint this last term may have a number of interpretations: failure to resolve a mistake that caused the horizon h implicit forecasting error, a new mistake introduced when the forecast is revised at horizon $h-1$, or both.

The mean square forecast revision between horizon h and $h-1$ is calculated as

$$\text{MSFR}_{t|h-1,h} \equiv \text{E}(d^2_{t|h-1,h}) \tag{3.29a}$$

$$= \sigma^2_{\omega_{h-1}} + \sigma^2_{\zeta_h} + \sigma^2_{\zeta_{h-1}}. \tag{3.29b}$$

where the positive sign on $\sigma^2_{\zeta_{h-1}}$ indicates that MSFR will always be greater than MSE as defined by (3.28c) when the horizon $h-1$ forecast contains implicit forecasting error. When a forecast revision can be characterised by (3.29b) all of the parameters lead to an increase in the variance of the forecast revision. Yet only the first two terms are associated with an improvement in forecast performance. This allows us to determine the horizon during which forecasts display the most inefficiency.

Based on the approach proposed by Isiklar and Lahiri (2007) we use the parameter $\sigma^2_{\zeta_{h-1}}$ in equations (3.28c) an (3.29b) to determine the inefficiency of the horizon $h-1$ forecast. Unlike the non-parametric forecast evaluation approach proposed by Isiklar and Lahiri (2007) our model uses a dynamic representation of the target variable and so we do not need to observe the target variable to obtain an estimate of MSE.

3.4.3 Forecast bias

Tests of internal consistency and forecast efficiency are based on the assumption that the forecasts are unbiased. Similarly, our model based approach to forecast evaluation assumes that the forecasts being evaluated are unbiased. In fact, in the taxonomy of forecast rationality tests, forecast unbiasedness is used as an initial stage in an increasingly more restrictive hierarchy of maintained hypothesis implied by rational expectations (Stekler, 2002; Clements, 2005). In this section we review extant tests of forecast unbiasedness and propose a state space approach for the detection and extraction of bias in multi-horizon forecasts.

The commonly used Mincer and Zarnowitz (1969) forecast rationality test is

essentially a test of forecast unbiasedness. The approach involves regressing the target variable on a constant and the forecast variable

$$y_t = \alpha_0 + \alpha_1 \hat{y}_{t|t-h} + \epsilon_t \tag{3.30}$$

and testing the joint null hypothesis $\alpha_0 = 0$ and $\alpha_1 = 1$. Tests are performed at individual h forecast horizons where rejection of the null is used to indicate forecast biasedness. One limitation of the Mincer and Zarnowitz (1969) test, as shown by Holden and Peel (1990), is that the null hypothesis is a sufficient but not necessary condition for forecast unbiasedness. The necessary and sufficient condition for forecast unbiasedness is $\alpha_0 = (1 - \alpha_1)\mathrm{E}(y_t)$. It is therefore possible for unbiased forecast series to fail this test. As an alternative Holden and Peel (1990) advocate testing whether forecasting error has a mean of zero. This can be done by regressing forecasting error on a constant, as follows

$$\hat{y}_{t|t-h} - y_t = e_{t|t-h} = \beta + \epsilon_t \tag{3.31}$$

where the null hypothesis of forecast unbiasedness involves testing the restriction $\beta = 0$.

Given the simplicity of this test, it is likely that forecasters would quickly become aware of any deviation from mean zero in their forecasting error. Forecasters could then take steps to improve their forecasting approach. Using this logic it is reasonable to assume that forecast bias would evolve over time. We may specify horizon specific forecasting error as an $I(2)$ process to capture a smooth trend using

$$e_{t|t-h} = \beta_{t|t-h} + \epsilon_t \qquad \epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2) \tag{3.32}$$

$$\beta_{t|t-h} = \beta_{t-1|t-1-h} + \gamma_{t-1} + \kappa_t \qquad \kappa_t \sim \mathrm{N}(0, \sigma_\kappa^2) \tag{3.33}$$

$$\gamma_t = \gamma_{t-1} + \psi_t \qquad \xi_t \sim \mathrm{N}(0, \sigma_\psi^2). \tag{3.34}$$

Once the trend deviation from zero in the forecasting error has been estimated we can remove the estimated value from the observed forecast series, as follows

$$\hat{y}^{\beta}_{t|t-h} = \hat{y}_{t|t-h} - \hat{\beta}_{t|t-h}. \tag{3.35}$$

A further limitation of the Mincer and Zarnowitz (1969) test concerns the assumptions made about the error term in the regression equation (3.30) (Holden and Peel, 1990; Miller, 1991). Specifically, the error term ϵ_t will take on the properties of the forecasting error. Were a forecaster to introduce measurement error into their forecast the forecast variable will be contemporaneously correlated with forecasting error. While this type of forecast will be unbiased the error term in the regression will be correlated with the forecasts variable, $\text{cov}(\hat{y}_{t|t-h}, \epsilon_t) \neq 0$, leading to downward bias in the coefficient α_1. This type of error can be captured by the implicit component in our model of multi-horizon rational-implicit forecasts.

To rule out contemporaneous correlation between forecasts and the error term in the Mincer and Zarnowitz (1969) regression test, we suggest subtracting an estimate of the implicit forecasting error component from our model from the bias corrected forecasts, as follows

$$\hat{y}^{*}_{t|t-h} = \hat{y}^{\beta}_{t|t-h} - \hat{\zeta}_{t|t-h}. \tag{3.36}$$

We then test the joint null hypothesis $\alpha^*_0 = 0$ and $\alpha^*_1 = 1$ in the following regresssion

$$y_t = \alpha^*_0 + \alpha^*_1 \hat{y}^{*}_{t|t-h} + \epsilon^{*}_{t|t-h}. \tag{3.37}$$

This serves as a check to see if we have successfully identified the multiple sources of forecasting error leading to a rejection of the forecast biasedness test in equation (3.30).

3.5 An illustrative example

The aim of this section is to present an empirical illustration of the forecast revision components that are obtained from our state space approach to multi-horizon forecast evaluation. Our empirical illustration uses a dataset containing meteorological weather forecasts and associated observations for Melbourne, Australia. Weather has an important bearing economic activity (Dell et al., 2014). One example is the relationship between weather and electricity demand (see Lee and Chiu (2011) for estimates in 24 OECD countries and Zhang et al. (2014) for estimates in China). Temperature is useful for explaining seasonality in electricity demand as temperature influences the demand for electricity to run heating and cooling appliances. Model based electricity demand forecasts commonly include temperature variables (Taylor and Buizza, 2003; Hong et al., 2015), and temperature variables have been used to forecast electricity supply (Zavala et al., 2009). While electricity demand is known to display diurnal patterns, at least in part related to weather, demand modelling based on weather variables are more useful at longer forecast horizons (Taylor, 2008). Ritter et al. (2011) showed that meteorological forecasts are important for explaining the price of temperature derivatives.

Over the last few decades, meteorological services in Australia have produced increasingly accurate weather forecasts, at ever longer forecast horizons (Stern, 2008; Stern and Davidson, 2015). Stern and Davidson (2015) evaluate weather forecasts over the period 1950 to 2014 by calculating the accuracy of meteorological forecasts relative to the accuracy of naïve forecasts; a measure known as forecast skill in this literature. Maximum daily temperature forecasts available during the mid-2000s, and produced 5-7 days out from the target date,

had forecast skill similar to forecasts available during the 1960s that were produced 1 day out from the target date. Similarly, forecasts available during 2014, that were produced 8-10 days out from the target date, had forecast skill similar to forecasts available during the mid-2000s that were produced 5-7 days out from the target date Stern and Davidson (2015). The aim of this empirical illustration is to show: how estimates of the rational and implicit components from our model of multi-horizon forecasts can be used to gain a deeper understanding of the sources of forecast performance.

3.5.1 Data

Our dataset contains experimental maximum daily temperature forecasts $\hat{y}_{t|t-h}$ and associated observations y_t for Melbourne, Australia available from http://www.weather-climate.com. The sample period we evaluate runs from 1 February 2009 to 31 December 2014 comprising a total of $t = 2160$ observations, with forecasts available at $h = 1, \ldots, 14$ days out from each observation date. Forecasts were produced in real-time using an experimental combined forecasting system approach as documented in Stern and Davidson (2015) and Stern (2006, 2007). Table 3.2 shows the combination of weights used in the combined forecasting system. A number of data sources are combined to produce the forecasts including official forecasts from the Australian Bureau of Meteorology, statistical forecasts, and climatology. The statistical forecasts essentially provide a local interpretation of output from a long-range global forecasting system. The combination of weights in the forecasting system change depending on the length of the forecast horizon h.

Figure 3.2 plots maximum daily temperature observations for Melbourne, Aus-

Table 3.2: Forecast combination weightings for maximum daily temperature in Melbourne, Australia.

Horizon	Official	Previous	Statistical	Climatology
$h = 1$ to 7 days	0.50	0.25	0.25	
$h = 8$ to 13 days		0.25	0.50	0.25
$h = 14$ days		0.50		0.50

Notes to Table 3.2: Stern and Davidson (2015) produce multi-horizon forecasts of maximum daily temperature using a forecast combination algorithm that draws on a number of data sources including official forecasts from the Australian Bureau of Meteorology, statistical forecasts (see Wilks (2011) for examples of statistical forecast methods for meteorological variables), and climatology. Composition of the forecast combination weightings change depending on the length of the forecast horizon.

tralia, (black line) from 1 February 2009 to 31 December 2014 and associated experimental forecasts produced $h = 7$ days (blue line) and $h = 14$ days (red line) out from the observation date. Table 3.3 reports descriptive statistics for several sub-periods of the dataset. The sub-periods of interest are summer and winter months, as well as periods before and after the upgrade of a Numerical Weather Prediction (NWP) model used as an input to the combined forecasting system. Southern hemisphere summer months are assumed to begin on September 21, roughly corresponding to the spring equinox, and southern hemisphere winter months assumed to begin on March 21, roughly corresponding with the autumn equinox.

All series in Figure 3.2 exhibit a clear seasonal pattern with warmer and more variable temperatures during the summer months and cooler and less variable temperatures during the winter months. The variance of summer temperatures are found to be statistically significantly higher than winter temperatures in both the observation and forecast series as shown by the F-test results in table 3.4. Stern and Davidson (2015) note that the competing influence of warm dry winds from the Australian interior, and cool moist winds from the Southern Ocean make forecasting temperature for Melbourne, Australia, particularly challenging. These competing influences are greatest during summer months. Summer forecasts are likely to have larger variability than winter forecasts

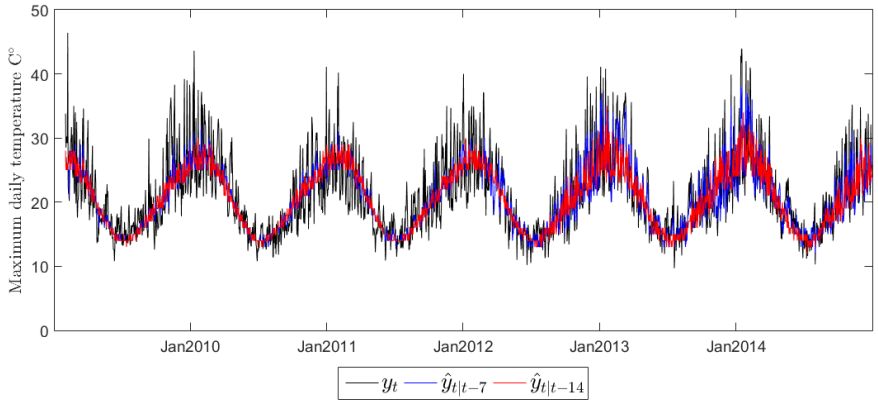


Figure 3.2: Maximum daily temperature observations and multi-horizon forecasts for Melbourne, Australia, 01-Feb-2009 to 31-Dec-2014.

Notes to Figure 3.2: Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. The figure plots maximum daily temperature observations (black line) and meteorological forecasts of maximum daily temperature produced at $h = 7$ days (blue line) and $h = 14$ days (red line) out from the observation date. Two notable features of this data are evident from this plot: 1) variability of observed maximum daily temperature is higher during the warmer months of the year compared to the cooler months of the year. A potential explanation for this relates to Melbourne's geographic location (Stern and Davidson, 2015). 2) variability of maximum daily temperature forecasts show a permanent increase begining around mid-2012. A potential explanation for this a major upgrade to one of the Numerical Weather Prediction models (implemented on 22 May, 2012) used as an input to these forecasts (Stern and Davidson, 2015). We use our multi-horizon forecast evaluation model to establish a relationship between these two data features and relative forecast performance over time.

as observed summer temperatures are more variable than observed winter temperatures.

As summer temperatures are more difficult to forecast, higher variability in summer forecasts, relative to winter forecasts, may be due to higher forecasting error variability caused by mistakes introduced by the forecaster. For instance, if the process driving summer temperature variability is not completely described by the forecasting system, increased forecast variability may be due to increased forecast error variance. Our aim is to show how estimates of the rational and implicit components of multi-horizon forecast revisions can be used to compare the information content of summer forecasts with the information content of winter forecasts.

The forecast series in Figure 3.2 exhibit a permanent increase in variability

Table 3.3: Descriptive statistics — maximum daily temperature observations and multi-horizon forecasts for Melbourne, Australia, 01-Feb-2009 to 31-Dec-2014.

	February 1, 2009 to May 21, 2012				May 22, 2012 to December 31, 2014			
	Summer		Winter		Summer		Winter	
	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.
$h = 14$	24.3	2.7	17.4	3.1	23.8	3.5	16.7	2.9
$h = 13$	24.3	2.8	17.5	3.1	24.0	3.8	16.8	3.0
$h = 12$	24.3	2.9	17.4	3.1	24.2	3.9	16.8	3.1
$h = 11$	24.3	2.9	17.4	3.0	24.2	4.0	16.8	3.2
$h = 10$	24.3	2.9	17.5	3.1	24.2	4.0	16.7	3.1
$h = 9$	24.3	2.9	17.5	3.1	24.1	3.9	16.7	3.0
$h = 8$	24.3	2.9	17.5	3.1	24.0	3.9	16.9	3.1
$h = 7$	24.3	3.0	17.5	3.2	24.4	4.6	17.1	3.3
$h = 6$	24.6	3.9	17.6	3.4	24.4	5.0	17.1	3.3
$h = 5$	24.7	4.3	17.7	3.5	24.4	5.2	17.2	3.4
$h = 4$	24.7	4.5	17.7	3.6	24.5	5.4	17.2	3.4
$h = 3$	24.6	4.7	17.7	3.7	24.6	5.5	17.2	3.4
$h = 2$	24.6	4.9	17.7	3.7	24.7	5.5	17.2	3.4
$h = 1$	24.7	5.0	17.7	3.7	24.6	5.6	17.2	3.4
$h = 0$	24.7	5.6	17.9	4.0	24.7	6.0	17.3	3.7

Notes to Table 3.3: Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. Over the period February 1, 2009 to May 21, 2012 forecast variability (standard deviation of $\hat{y}_{t|t-h}$) is monotonically decreasing in the forecast horizon as required by internal consistency. Over the period May 22, 2012 to December 31, 2014 forecast variability is not monotonically decreasing in the forecast horizon. Violation of this internal consistency property occurs at forecast horizons beyond $h = 7$ days.

beginning around mid-2012. Table 3.3 shows that the variability of the forecast series is higher after 22 May, 2012, while the variability of the observed series remains unchanged. Notationally the observed series is defined as $h = 0$ in Table 3.3. This finding is supported by F-test results in table 3.4 which show no difference in observed maximum daily temperature over the series, yet a statistically significant increase in the variance of the forecast after 22 May, 2012, for horizons $h = 2$ through $h = 13$. Again the observed series is defined as $h = 0$. In their evaluation of the forecast series Stern and Davidson (2015) describe an increase in forecast skill beginning mid-2012 which they attribute to the major upgrade of a global NWP model (implemented on 22 May, 2012). Output from this NWP model is used as an input to the statistical forecasts listed in Table 3.2.

Table 3.4: Differences in the variability of the maximum daily temperature observation and forecast series over time.

	Null hypothesis: $\text{Var}(\hat{y}_{t\|t-h})$ Winter $\geq$ $\text{Var}(\hat{y}_{t\|t-h})$ Summer	**Null hypothesis:** $\text{Var}(\hat{y}_{t\|t-h})$ Feb 1, 2009 to May 21, 2012 $\geq$ $\text{Var}(\hat{y}_{t\|t-h})$ May 22, 2012 to Dec 31, 2014
	F test statistic	F test statistic
$h = 14$	0.969	0.907
$h = 13$	0.875*	0.837*
$h = 12$	0.838*	0.804*
$h = 11$	0.835*	0.783*
$h = 10$	0.828*	0.786*
$h = 9$	0.826*	0.828*
$h = 8$	0.847*	0.833*
$h = 7$	0.733*	0.748*
$h = 6$	0.590*	0.841*
$h = 5$	0.542*	0.864*
$h = 4$	0.517*	0.877*
$h = 3$	0.499*	0.878*
$h = 2$	0.484*	0.897
$h = 1$	0.465*	0.915
$h = 0$	0.454*	0.935

Notes to Table 3.4: Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. * Indicates a rejection of the Null hypothesis at $\alpha = 0.05$.

Figure 3.3 reproduces the skill measure reported in Stern and Davidson (2015) based on a moving average of the percentage of variance in the observations explained by the forecasts. This skill measure is equivalent to the R^2 obtained when observed temperature deviations from the climatological mean are regressed on forecast temperature deviations from the climatological mean. A clear upward trend in the R^2 skill metric is evident during the 2012 period for all the forecast horizons in Figure 3.3. The skill metric stabilises at the higher level during early 2013 and remains at that level for the remainder of the sample period.

Figure 3.4 plots forecast root mean square error (RMSE) over the same period. Plots for the average of horizons $h = 1$ to $h = 4$ and $h = 5$ to $h = 7$ show a

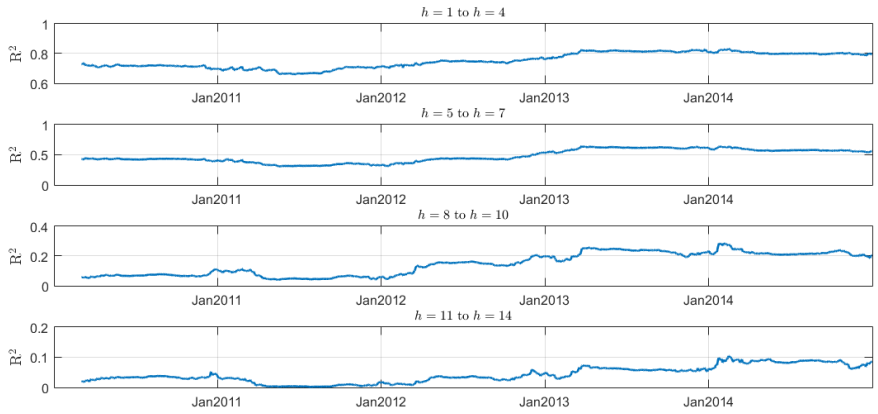


Figure 3.3: Percentage of variance explained by multi-horizon temperature forecasts for Melbourne, Australia, 01-Feb-2009 to 31-Dec-2014.

Notes to Figure 3.3: Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. The skill metric reported in this chart is the 365 day moving average of R^2 when observed temperature deviations from the climatological mean are regressed on forecast temperature deviations from the climatological mean.

reduction in RMSE during the 2012 period that supports the increase in skill evident in Figure 3.3. There is a slight increase in RMSE toward the end of the sample period for these horizons, but the new RMSE level is clearly lower than the RMSE level in the pre-2012 sample period. Contrast this finding with the RMSE plots for horizons $h = 8$ to $h = 10$ and $h = 11$ to $h = 14$ which do not support the story told by the associated skill metric plots in Figure 3.3. At longer forecast horizons, RMSE is stable over the 2012 period, before increasing to a higher level during early 2013, and remaining at that new level for the remainder of the sample period. Despite forecasts at these longer horizons exhibiting a higher level of skill after the NWP model upgrade, RMSE, which is arguably a more useful metric for the end users of such forecasts, suggests that forecast performance worsens after the NWP upgrade.

One aim of this empirical illustration, is to show how estimates of the rational and implicit components of multi-horizon forecast revisions can be used to

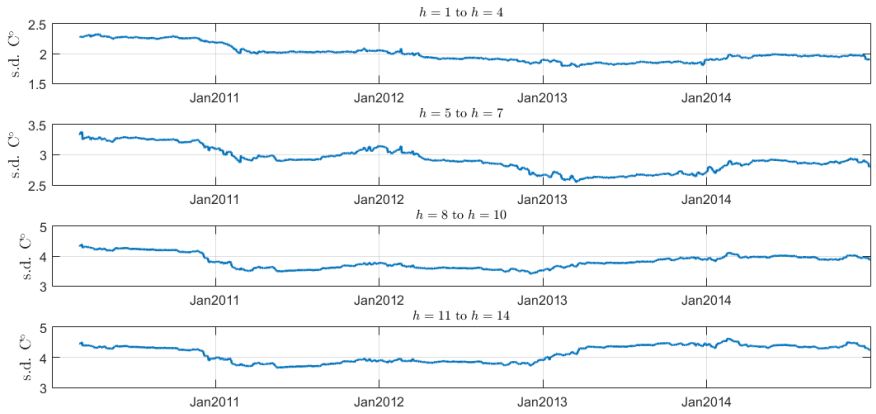


Figure 3.4: RMSE of multi-horizon temperature forecasts for Melbourne, Australia, 01-Feb-2009 to 31-Dec-2014.

Notes to Figure 3.4: Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. The forecast performance metric reported in this chart is the 365 day moving average of Root Mean Square Forecasting Error (RMSE).

compare the information content of forecasts produced before and after the NWP model upgrade. These descriptive statistics suggest a number of hypothesis which are testable in the framework of this chapter: 1) does information accumulation differ over forecast horizons and seasons; 2) does the variation in implicit forecasting error differ across horizons or across seasons; and 3) does the NWP model upgrade have an impact on the structure of multi-horizon forecasting error.

3.5.2 Results

In this subsection we evaluate maximum daily temperature forecasts over the period 1 February, 2009, to 31 December, 2014. We are interested in how the forecasts perform over the entire sample period, and within a number of sub-periods during the sample.

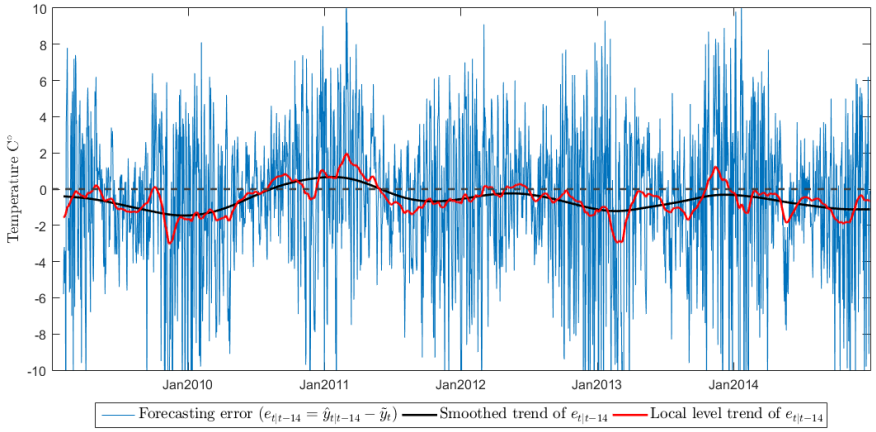
Our model of multi-horizon forecasts assumes that the forecasts being evalu-

ated are unbiased. Figure 3.5a–3.5c shows estimates of bias components $\beta_{t|t-h}$ estimated using equations (3.32), (3.33), (3.34). The top panel plots forecasting error along with Kalman filtered smoothed estimates of forecast bias. Estimates are obtained using a local linear trend model and an integrated random walk model. The local linear trend model is preferred by the AIC model selection criteria. The middle panel plots bias estimates from the preferred local linear trend model for forecasts at $h = 1$ day out to $h = 7$ days from the target. The bias estimates are stable over time, with higher bias associated with longer forecast horizons. This finding is evidenced by the non zero values associated with the bias estimates for h_7 in panel (b) of Figure 3.5. The bottom panel plots bias estimates for forecasts at $h = 8$ days out to $h = 14$ days from the target. The bias in longer horizon forecasts displays signs of an irregular cyclical pattern. Before proceeding, we use the estimates of $\beta_{t|t-h}$ to extract bias from each $\hat{y}_{t|t-h}$ series using equation (3.35), as described in Section 3.4.

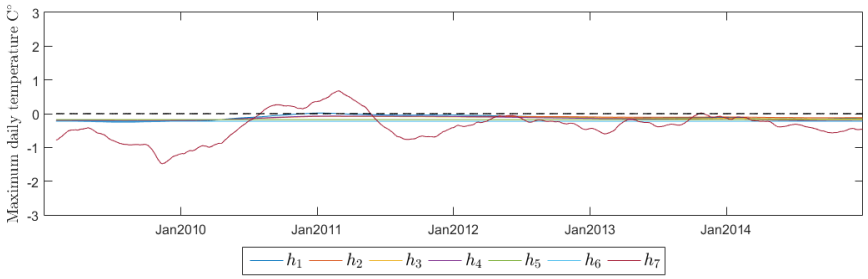
We fit two model specifications to the data: a rational model, given by equation (3.19) and equation (3.20); and a rational-implicit model, given by equation (3.17) and equation (3.18). The rational model assumes that the only source of forecasting error is unforecastble information realised after the forecasts are produced. The rational-implicit model assumes that forecasters make horizon specific mistakes by introducing information unrelated to the target variable into their forecasts. We determine which of these model specifications best describes the data using the Akaike information criterion (AIC).

Full sample

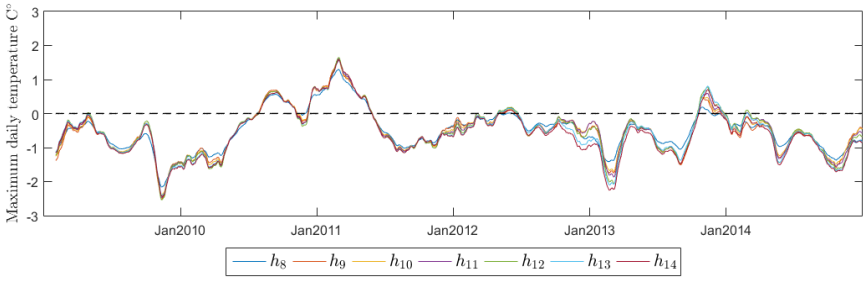
Table 3.5 shows parameter estimates from the rational and rational-implicit model when they are fit to bias corrected maximum daily temperature forecasts from horizons $h = 1$ through $h = 14$. The rational forecast revision parameter estimates σ_{ω_h} in Table 3.5 report the standard deviation of the rational component of forecast revisions between each adjacent forecast horizon. As a rational forecast revision updates the forecast from the previous horizon with new information, we describe the parameter σ_{ω_h} as an estimate of the increase in information content owing to the forecast revision. Parameter estimate σ_{ω_0} can be thought of as a measure of the forecaster's ignorance as it estimates the information that becomes available when the target variable is realised. The implicit forecasting error parameter estimates σ_{ζ_h}, in the rational-implicit model, report the standard deviation of the non-rational component of forecasting error. As described in Section 3.2.3 implicit forecasting error is a mistake introduced into the forecast that is unrelated to the target variable.



(a)



(b)



(c)

Figure 3.5: State estimates of bias in multi-horizon forecasts of maximum daily for Melbourne, Australia.

Notes to Figure 3.5: Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. a) State component estimates from a local linear trend model (red line) and an integrated random walk model (black line) applied to horizon $h = 14$ forecasting error $et|t-14 = \hat{y}_{t|t-14} - y_t$ (blue line). b) State component estimates from a local linear trend model applied to forecasting error $et|t-h = \hat{y}_{t|t-h} - y_t$ at horizons $h = 1$ to $h = 7$. c) State component estimates from a local linear trend model applied to forecasting error $et|t-h = \hat{y}_{t|t-h} - y_t$ at horizons $h = 8$ to $h = 14$.

Information criteria indicate that the multi-horizon maximum daily temperature forecasts are best characterised by the rational-implicit model. For both models in Table 3.5 the rational forecast revision parameter estimates at short horizons (σ_{ω_0} to σ_{ω_6}) are virtually identical, and large relative to their standard errors. The forecast revision between horizon $h = 7$ and horizon $h = 6$ adds the largest amount of new information into the forecasts. As the forecast horizon becomes shorter the marginal increase in forecast information content declines. This has implications for end users of the forecasts who make decisions based on future weather events and face a trade-off between information content and timeliness. Consider a forecast user who orders goods that have weather sensitive demand. For instance, an ice-cream vendor may face a trade-off between ordering in bulk at discounted prices based on uncertain long horizon weather forecasts, or placing smaller orders each day at standard prices based on short horizon forecasts containing more certainty.

At longer forecast horizons the rational and rational-implicit models differ in their parameter estimates for σ_{ω_7} to $\sigma_{\omega_{13}}$. The rational-implicit model is more flexible as it allows for the possibility that mistakes enter the forecast revision process. Forecasting mistakes, or implicit forecasting error, lead to an increase in forecast variability without a subsequent increase in forecast information content. In the preferred rational-implicit model σ_{ω_h} estimates indicate that the forecast revisions contain increasingly smaller amounts of information as the forecast horizon increases beyond $h = 7$. This has intuitive appeal as we would expect there to be less relevant information that forecasters could incorporate into forecasts at long horizons. The incorrectly specified rational model suggest that forecast revisions contain increasingly larger amounts of information as the forecast horizon increases.

Table 3.5: The information content of multi-horizon forecasts of maximum daily temperature for Melbourne, Australia between February 1, 2009, and December 31, 2014.

		Alternative models			
		Rational		**Rational + Implicit**	
Parameters		Estimate	Std err	Estimate	Std err
Rational revision	$\sigma_{\omega_{13}}$	1.57	(0.04)	0.05	(0.97)
	$\sigma_{\omega_{12}}$	1.65	(0.04)	0.82	(0.03)
	$\sigma_{\omega_{11}}$	1.56	(0.04)	0.63	(0.03)
	$\sigma_{\omega_{10}}$	1.53	(0.04)	0.65	(0.03)
	σ_{ω_9}	1.47	(0.04)	0.68	(0.03)
	σ_{ω_8}	1.42	(0.04)	0.99	(0.02)
	σ_{ω_7}	1.42	(0.04)	1.31	(0.02)
	σ_{ω_6}	1.56	(0.04)	1.56	(0.02)
	σ_{ω_5}	1.09	(0.04)	1.09	(0.02)
	σ_{ω_4}	0.97	(0.02)	0.97	(0.01)
	σ_{ω_3}	0.82	(0.02)	0.82	(0.01)
	σ_{ω_2}	0.73	(0.02)	0.73	(0.01)
	σ_{ω_1}	0.70	(0.02)	0.70	(0.01)
	σ_{ω_0}	1.69	(0.03)	1.69	(0.02)
Implicit error	$\sigma_{\zeta_{14}}$			1.19	(0.04)
	$\sigma_{\zeta_{13}}$			1.06	(0.02)
	$\sigma_{\zeta_{12}}$			1.05	(0.02)
	$\sigma_{\zeta_{11}}$			1.03	(0.02)
	$\sigma_{\zeta_{10}}$			1.01	(0.02)
	σ_{ζ_9}			0.88	(0.02)
	σ_{ζ_8}			0.54	(0.02)
	σ_{ζ_7}			0.03	(0.75)
	σ_{ζ_6}			0.02	(0.73)
	σ_{ζ_5}			0.01	(0.70)
	σ_{ζ_4}			0.01	(0.45)
	σ_{ζ_3}			0.01	(0.47)
	σ_{ζ_2}			0.01	(0.44)
	σ_{ζ_1}			0.02	(0.55)
Akaike Info Criterion		111,388		109,330	
Logarithmic likelihood		-55,678		-54,636	

Notes to Table 3.5: Standard errors are in parentheses. Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. The forecasts $\hat{y}_{t|t-h}$ are bias adjusted, where the bias follows a local linear trend process. The two alternative models of multi-horizon forecasts we consider are: rational ($\hat{y}^{\beta}_{t|t-h} = \tilde{y}_t + \nu_{t|t-h}$); and rational-implicit ($\hat{y}^{\beta}_{t|t-h} = \tilde{y}_t + \nu_{t|t-h} + \zeta_{t|t-h}$). Comparing AIC values for these three models we conclude that the data is best characterised by the rational-implicit model. Parameter $\sigma_{\omega_{h|t-h}}$ is the standard deviation of the rational forecast revision between horizon $h+1$ and horizon h. Parameter σ_{ζ_h} is the standard deviation of the implicit forecasting error component at horizon h. Parameter estimates for the rational-implicit model show a peak in the variability of rational forecast revisions at horizon $h = 6$. We interpret this parameter as the increase in information content due to forecast revision. Parameter estimates σ_{ω_5} to σ_{ω_1} suggest that the rate of increase in information content declines over horizons $h = 5$ to $h = 1$. The σ_{ω_0} parameter estimate shows the increase in information content when the target variable is realised. Parameter estimates for the rational-implicit model show that the forecaster introduces implicit error into forecasts at horizon $h = 7$ to horizon $h = 14$. We interpret parameters σ_{ζ_h} as the inefficient use of information by forecasters.

Parameter estimates σ_{ζ_h} help explain why the parameter estimates σ_{ω_h} differ between the rational and rational-implicit models. Table 3.5 indicates that forecast variability is relatively stable between horizons $h = 8$ and $h = 14$. We expect forecast variability to decline at longer horizons as less information is available to incorporate into the forecasts. Parameter estimates σ_{ζ_h} in the rational implicit-model increase in h, suggesting that the variability in long horizon forecasts is due to the introduction of implicit forecasting error, rather than relevant information content. This is supported by the decline in parameter estimates σ_{ω_h} as h increases.

Figure 3.6 shows the mean square forecast revisions $\text{MSFR}_{t|h,h-1}$, and their components, when forecast revisions are produced from 13 days to one day out from the target date. The variability of the forecast revision $d_{t|h,h+1}$ at each h is the combined effect of the rational forecast revision component σ_{ω_h} and two implicit forecasting error components $\sigma_{\zeta_{h+1}}$ and σ_{ζ_h}. As discussed in Section 3.4.2 $\sigma_{\zeta_h} > 0$ represents forecasts inefficiency at horizon h. Figure 3.6 indicates that the forecasts produced at horizons $h = 13$ to $h = 8$ are inefficient.

We now compare the results of our model based approach to forecast evaluation with existing tests of forecast rationality. We begin by exploring visually, some properties we expect rational multi-horizon forecasts to possess. The relationship $\text{Var}(y_t) = \text{Var}(\hat{y}^*_{t|t-h}) + \text{Var}(e^*_{t|t-h}) \geq \text{Var}(\hat{y}^*_{t|t-h})$ is implied by the rational expectations hypothesis. The direction of implication means that this property is a necessary but not sufficient condition for forecast rationality. Forecasts that possess this property may be described as internally consistent, but not necessarily rational. Figure 3.7 suggests that the forecasts in our dataset are internally consistent. The property that $\text{Var}(\hat{y}_{t|t-h}) \leq \text{Var}(y_t)$ is

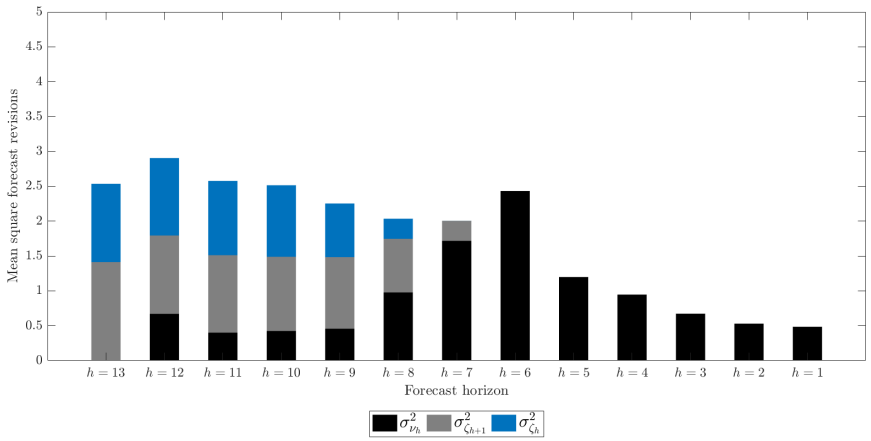


Figure 3.6: The components of mean square forecast revisions in maximum daily temperature forecasts.

Notes to Figure 3.6: The multi-horizon forecasts of maximum daily temperature produced by Stern and Davidson (2015) are best characterised by the rational-implicit model ($\hat{y}^{\beta}_{t|t-h} = \tilde{y}_t + \nu_{t|t-h} + \zeta_{t|t-h}$). See Table 3.5 for parameter estimates, their standard errors, and AIC values for alternative models. Parameter $\sigma_{\omega_{h|t-h}}$ is the standard deviation of the rational forecast revision between horizon $h+1$ and horizon h. Parameter σ_{ζ_h} is the standard deviation of the implicit forecasting error component at horizon h. We interpret parameters σ_{ζ_h} as the inefficient use of information by forecasters.

satisfied at all h horizons. The property that $\text{Var}(\hat{y}_{t|t-h})$ is non-increasing in h appears to be satisfied, with the possible exception of $h = 9$ and $h = 10$. Forecasting error variability, measured as mean square forecasting error (MSFE_t), satisfies the property that $\text{Var}(e^*_{t|t-h})$ is non-decreasing in h.

Table 3.6 presents the results of formal variance and covariance bounds tests implied by forecast rationality, as proposed by Patton and Timmermann (2012). The null hypothesis of forecast rationality is not rejected in any these tests. The results presented here provide an example of how forecasts can be inefficient and yet still have variance and covariance bounds that satisfy the monotonic patterns required by Patton and Timmermann (2012) tests.

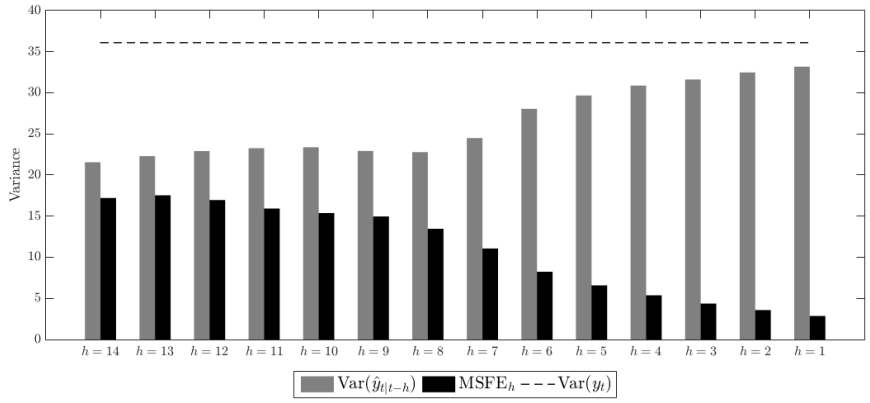


Figure 3.7: Mean squared errors and forecast variances suggest that the multi-horizon maximum daily temperature forecasts are internally consistent.

Notes to Figure 3.7: Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. The following internally consistent properties are implied by forecast rationality: $\text{Var}(\hat{y}_{t|t-s}) \geq \text{Var}(\hat{y}_{t|t-l})$ for $l > s$; $\text{Var}(e_{t|t-s}) \leq \text{Var}(e_{t|t-l})$ for $l > s$; and $\text{Var}(y_t) \geq \text{Var}(\hat{y}_{t|t-l})$ for all h. This chart suggests that the maximum daily temperature forecasts are internally consistent. Table 3.6 reports the results of Patton and Timmermann (2012) tests for these properties.

Table 3.6: Results of internal consistency tests proposed by Patton and Timmermann (2012).

	Horizons	
Test	$h = 1$ to $h = 7$	$h = 8$ to $h = 14$
$\text{Var}(\hat{y}_{t\|t-s}) \geq \text{Var}(\hat{y}_{t\|t-l})$	0.96	0.85
$\text{Var}(e_{t\|t-s}) \leq \text{Var}(e_{t\|t-l})$	0.95	0.85
$\text{Var}(d_{t\|s,m}) \leq \text{Var}(d_{t\|s,l})$	0.91	0.95
$\text{Cov}(\hat{y}_{t\|t-s}, \tilde{y}_t) \geq \text{Cov}(\hat{y}_{t\|t-l}, \tilde{y}_t)$	0.95	0.93
$\text{Cov}(\hat{y}_{t\|t-m}, \hat{y}_{t\|t-s}) \geq \text{Cov}(\hat{y}_{t\|t-l}, \hat{y}_{t\|t-s})$	0.93	0.95
$\text{Var}(d_{t\|s,l}) \leq 2\text{Cov}(\tilde{y}_t, d_{t\|s,l})$	0.78	0.50
$\text{Var}(d_{t\|m,l}) \leq 2\text{Cov}(\hat{y}_{t\|t-s}, d_{t\|m,l})$	0.77	0.13

Notes to Table 3.6: Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. Values reported in Table 3.6 are p-values corresponding with the Patton and Timmermann (2012) tests PT1 to PT10 in Table (3.1). The null hypothesis of forecast rationality is not rejected by any of these tests. The inequality tests, based on Gourieroux et al. (1982) and Wolak (1987, 1989) cannot handle dimensions greater than 10. We use Matlab code for the inequality tests provided by Andew Patton and so partition the forecast horizon into $h = 1$ to $h = 7$ and $h = 8$ to $h = 14$ and test each partition separately.

When evaluating forecast rationality, tests of forecast biasedness are usually the first step in a series of maintained hypothesis. In fact the tests proposed by Patton and Timmermann (2012) and the approach we outline in this chapter assume that the forecasts are unbiased. We have chosen to present the results of forecast biasedness test at the end of this section so that we can illustrate how implicit forecasting error impacts the test results.

Coefficient estimates from the commonly used regression tests of forecast biasedness, reported in the first panel of Table 3.7 indicate that the assumption of unbiasedness, as required for Patton and Timmermann (2012) rationality tests, is invalid for the observed forecast series. An F-test the joint null hypothesis that the coefficients of Equation 3.30 satisfy $\alpha_0 = 0$ and $\alpha_1 = 1$ is rejected at conventional levels of significance, for all forecast horizons.

To determine if horizon specific bias is to blame for this rejection we re-run the tests using our bias corrected forecasts. The second panel of Table 3.7 indicates that bias corrected forecasts for all but two forecast horizons do not reject the joint null hypothesis $\alpha_0 = 0$ and $\alpha_1 = 1$. Contemporaneous correlation of the forecast variable and the regression error term, as occurs when implicit forecasting error is present, can lead to downward bias in the coefficient α_1.

To check whether the downward bias in α_1^{β} is due to the presence of implicit forecasting error we remove an estimate of implicit forecasting error from the bias corrected forecasts using the method described in Section 4.3. The third panel of Table 3.7 indicates that the bias and implicit error corrected forecasts (rational forecasts) do not reject the joint null hypothesis $\alpha_0^* = 0$ and $\alpha_1^* = 1$ at any forecast horizon, using equation 3.37.

Table 3.7: Mincer and Zarnowitz (1969) tests of forecast biasedness of actual forecasts and corrected forecasts.

	Original forecasts					Bias corrected forecasts					Rational forecasts				
	α_0	t-stat.	α_1	t-stat.	p-val.	α_0^β	t-stat.	α_1^β	t-stat.	p-val.	α_0^*	t-stat.	α_1^*	t-stat.	p-val.
$h = 14$	1.70	(3.57)	0.95	(-2.12)	0.00	0.72	(1.56)	0.97	(-1.29)	0.13	-0.28	(-0.66)	1.01	(0.59)	0.78
$h = 13$	2.13	(4.50)	0.92	(-3.07)	0.00	1.17	(2.56)	0.95	(-2.30)	0.02	-0.28	(-0.66)	1.01	(0.59)	0.78
$h = 12$	2.09	(4.67)	0.92	(-3.20)	0.00	1.08	(2.50)	0.95	(-2.25)	0.03	-0.43	(-1.00)	1.02	(0.90)	0.55
$h = 11$	1.76	(3.94)	0.94	(-2.58)	0.00	0.80	(1.85)	0.96	(-1.69)	0.15	-0.53	(-1.24)	1.02	(1.12)	0.40
$h = 10$	1.56	(3.50)	0.95	(-2.19)	0.00	0.64	(1.48)	0.97	(-1.36)	0.30	-0.62	(-1.48)	1.03	(1.34)	0.28
$h = 9$	1.20	(2.74)	0.97	(-1.43)	0.00	0.26	(0.62)	0.99	(-0.53)	0.77	-0.71	(-1.71)	1.03	(1.55)	0.18
$h = 8$	0.46	(1.08)	1.00	(0.07)	0.00	-0.29	(-0.71)	1.01	(0.68)	0.77	-0.68	(-1.68)	1.03	(1.53)	0.20
$h = 7$	0.08	(0.23)	1.01	(0.70)	0.00	-0.37	(-1.04)	1.02	(0.95)	0.55	-0.37	(-1.04)	1.02	(0.95)	0.55
$h = 6$	0.28	(0.93)	1.00	(-0.17)	0.00	0.06	(0.18)	1.00	(-0.17)	0.98	0.06	(0.18)	1.00	(-0.17)	0.98
$h = 5$	0.22	(0.83)	1.00	(-0.16)	0.00	0.05	(0.17)	1.00	(-0.16)	0.98	0.05	(0.17)	1.00	(-0.16)	0.98
$h = 4$	0.19	(0.84)	1.00	(-0.18)	0.01	0.04	(0.19)	1.00	(-0.17)	0.98	0.04	(0.18)	1.00	(-0.17)	0.98
$h = 3$	0.09	(0.46)	1.00	(0.19)	0.01	-0.05	(-0.22)	1.00	(0.21)	0.97	-0.05	(-0.22)	1.00	(0.21)	0.97
$h = 2$	0.10	(0.55)	1.00	(0.10)	0.01	-0.03	(-0.14)	1.00	(0.13)	0.99	-0.03	(-0.14)	1.00	(0.13)	0.99
$h = 1$	0.09	(0.62)	1.00	(0.16)	0.01	-0.03	(-0.23)	1.00	(0.21)	0.97	-0.03	(-0.23)	1.00	(0.22)	0.97

Notes to Table 3.7: Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. t-statistics are in parentheses. t-Stat on α_1 compares estimate to 1 not 0. The decision rule is to reject the joint null hypothesis that $\alpha_0 = 0$ and $\alpha_1 = 1$ if the p-value is less than 0.05.

To summarise, this subsection has illustrated how our approach can be used to detect multiple sources of forecast sub-optimality. The experimental multi-horizon forecasts of maximum daily temperature for Melbourne, Australia were found to contain horizon specific forecast bias and we found evidence that forecasters introduce mistakes into the forecasts produced at horizons longer than seven days. In the subsections that follow we evaluate maximum daily temperature forecast performance during a number of sub-periods of the sample. We focus on longer horizon forecasts to reduce the number of parameters that need to be estimated.

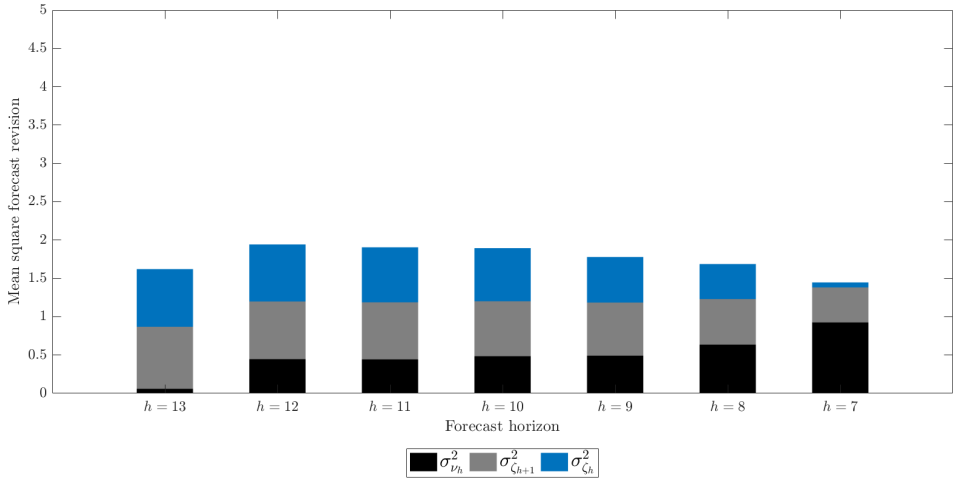
Forecast efficiency before and after the NWP model upgrade

The difference in forecast variability that is evident in Figure 3.2 before and after May 22, 2012 may be associated with changes in the structure of forecasting errors.

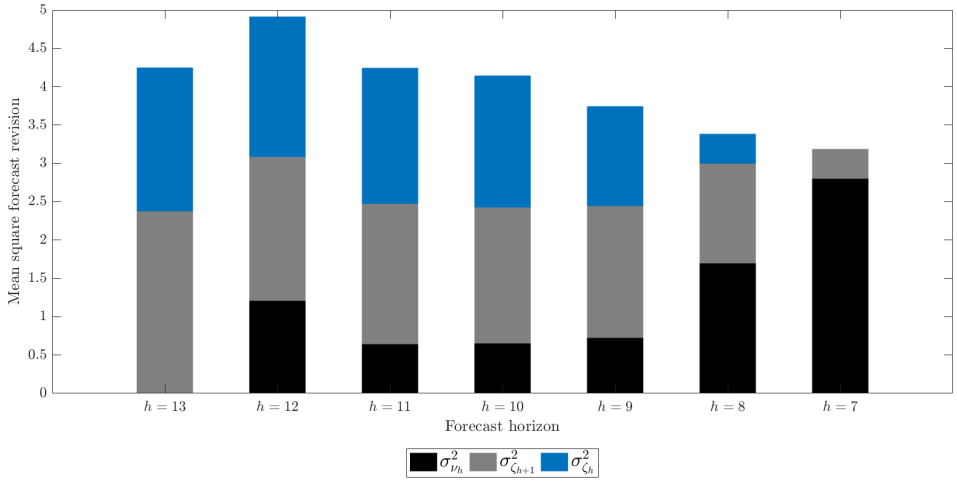
We use a rational-implicit model that contains a different set of parameters before and after 22 May, 2012, to evaluate the multi-horizon maximum daily temperature forecasts. See the top left panel Table B.1 for parameter estimates and their standard errors. Figure 3.8a and 3.8b illustrate that forecast revisions contain more relevant information and more inefficiency following the NWP model upgrade.

Forecast efficiency during summer and winter

Maximum daily temperature observations and forecasts show increased variability during the summer months. Higher summer temperature variability



(a)

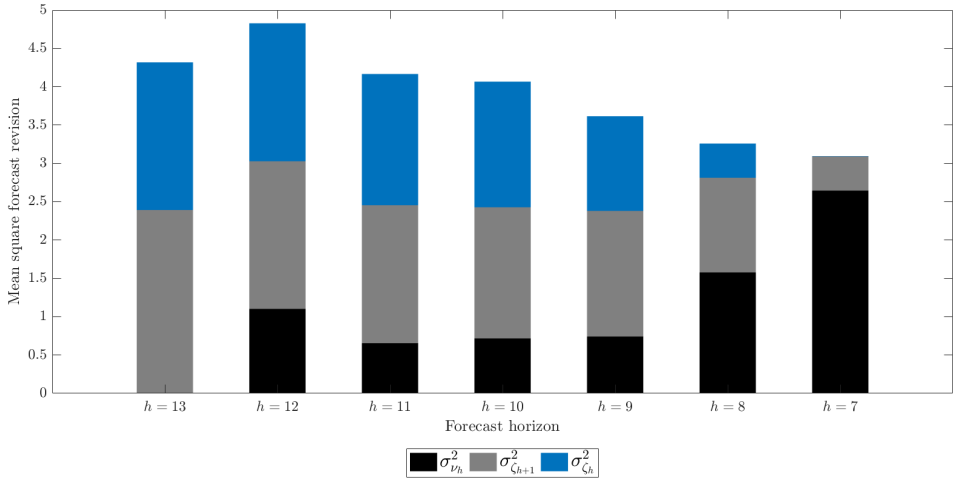


(b)

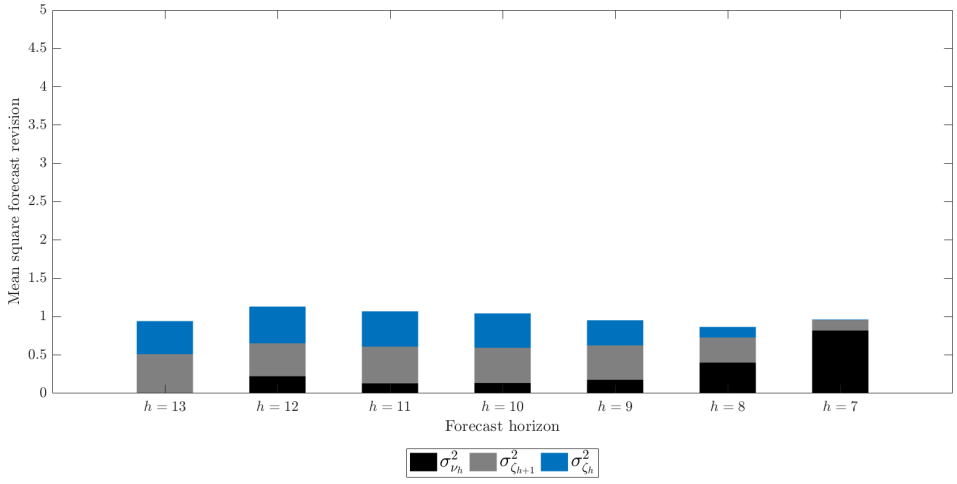
Figure 3.8: MSFR and forecasting error components before and after the NWP model upgrade.

Notes to Figure 3.8: The multi-horizon forecasts of maximum daily temperature produced by Stern and Davidson (2015) are best characterised by the rational-implicit model ($\hat{y}^{\beta}_{t|t-h} = \tilde{y}_t + \nu_{t|t-h} + \zeta_{t|t-h}$). See Table B.1 for parameter estimates, their standard errors, and AIC values for alternative models. Parameter $\sigma_{\omega_{h|t-h}}$ is the standard deviation of the rational forecast revision between horizon $h+1$ and horizon h. Parameter σ_{ζ_h} is the standard deviation of the implicit forecasting error component at horizon h. We interpret parameters σ_{ζ_h} as the inefficient use of information by forecasters. The information content of the multi-horizon forecasts, represented by σ_{ω}, is lower during the period before the 22 May, 2012, upgrade of the Numerical Weather Prediction (NWP) model a) than after the NWP model upgrade b). Implicit forecasting error is also higher after 22 May, 2012.

may lead to more information content in the summer forecasts relative to winter, more forecasting mistakes, or both.



(a)



(b)

Figure 3.9: MSFR and forecasting error components for summer and winter months.

Notes to Figure 3.9: The multi-horizon forecasts of maximum daily temperature produced by Stern and Davidson (2015) are best characterised by the rational-implicit model ($\hat{y}^{\beta}_{t|t-h} = \tilde{y}_t + \nu_{t|t-h} + \zeta_{t|t-h}$). See Table B.1 for parameter estimates, their standard errors, and AIC values for alternative models. Parameter $\sigma_{\omega_{h|t-h}}$ is the standard deviation of the rational forecast revision between horizon $h+1$ and horizon h. Parameter σ_{ζ_h} is the standard deviation of the implicit forecasting error component at horizon h. We interpret parameters σ_{ζ_h} as the inefficient use of information by forecasters. The information content of the multi-horizon forecasts, represented by σ_{ω}, is lower during the winter months a) than during the summer months b). Implicit forecasting error is also higher during the winter months.

We use a rational-implicit model that contains a different set of parameters for summer and winter months to evaluate the multi-horizon maximum daily temperature forecasts. See the top right panel of Table B.1 for parameter es-

timates and their standard errors. Figure 3.9a and 3.9b illustrate that there is higher information content and more forecasting mistakes in forecast revisions during the summer months, relative to the winter months. This higher information content reflects a higher variability in the observed series that needs to be incorporated in to the forecasts for them to perform adequately.

Forecast efficiency during summer and winter, before and after the NWP model upgrade

We now evaluate forecasts during summer and winter months, before and after the NWP model update. See the bottom panel of Table B.1 for parameter estimates and their standard errors. Comparing AIC values for the three models in Table B.1 we conclude that the data is best characterised by the rational-implicit model with separate parameters for summer months and winter months before and after May 22, 2012.

Figures 3.10a, 3.10b, 3.11a, and 3.11b illustrate the results of the preferred model. Forecasts produced during both summer and winter months show a higher rate of information accumulation after an upgrade to the NWP model on May 22, 2012, relative to the period before this upgrade occurred. Forecasting mistakes also increase after May 22, 2012.

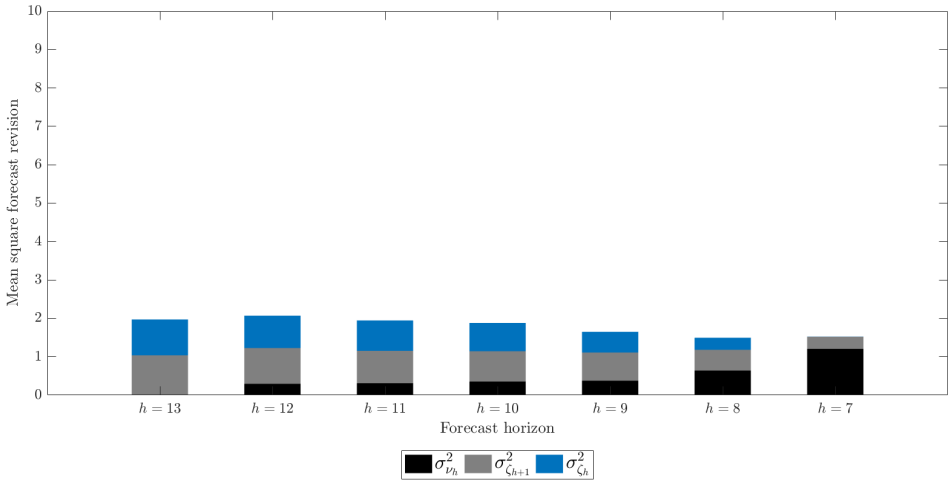
Table 3.8 reports the ratio of the standard deviation of implicit forecasting error to root mean square error (RMSE) for each of the forecast horizons evaluated. Comparing this ratio across time shows that it is possible for total forecasting error to decline even though the size of forecasting mistakes increase. In most cases however, both implicit forecasting error and RMSE increased after the NWP model upgrade. Another feature of interest is that differences

in forecast efficiency between summer and winter were largely absent prior to the NWP upgrade. Following the upgrade summer forecast revisions are estimated to contain a larger amount of forecasting mistakes.

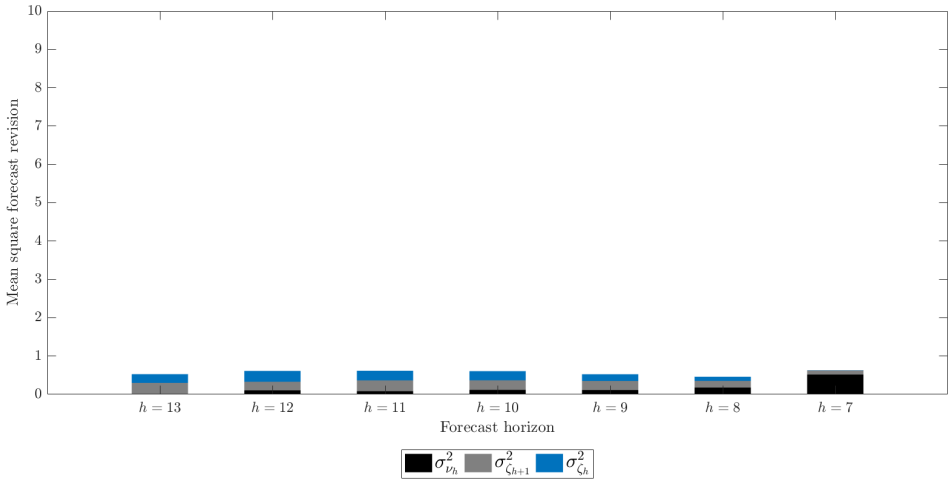
Table 3.8: The efficiency comparison between sub-samples of pre and post NWP model upgrade and the subsamples of winter and summer seasons.

	Feb 1, 2009 to May 21, 2012					
	Summer			**Winter**		
	σ_{ζ_h}	RMSE	Ratio	σ_{ζ_h}	RMSE	Ratio
h=14	1.01	4.91	0.21	0.55	2.76	0.20
h=13	0.97	5.03	0.19	0.47	2.71	0.18
h=12	0.92	5.01	0.18	0.53	2.72	0.20
h=11	0.89	4.96	0.18	0.50	2.70	0.19
h=10	0.86	4.87	0.18	0.49	2.65	0.18
h=9	0.74	4.82	0.15	0.42	2.60	0.16
h=8	0.56	4.67	0.12	0.33	2.51	0.13
h=7	0.07	4.42	0.02	0.06	2.37	0.03
	May 22, 2012 to Dec 31, 2014					
	Summer			**Winter**		
	σ_{ζ_h}	RMSE	Ratio	σ_{ζ_h}	RMSE	Ratio
h=14	2.04	5.53	0.37	0.88	2.92	0.30
h=13	1.83	5.55	0.33	0.83	2.97	0.28
h=12	1.76	5.33	0.33	0.85	2.93	0.29
h=11	1.71	5.00	0.34	0.85	2.80	0.30
h=10	1.67	4.90	0.34	0.84	2.80	0.30
h=9	1.46	4.86	0.30	0.72	2.68	0.27
h=8	0.79	4.44	0.18	0.42	2.57	0.16
h=7	0.13	3.69	0.03	0.12	2.28	0.05

Notes to Table 3.8: Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date.



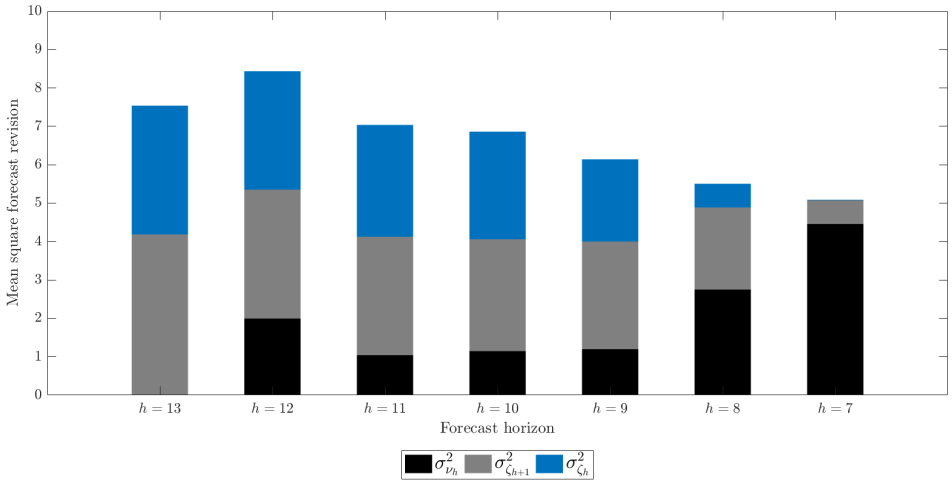
(a) Summer, before 22 May, 2012



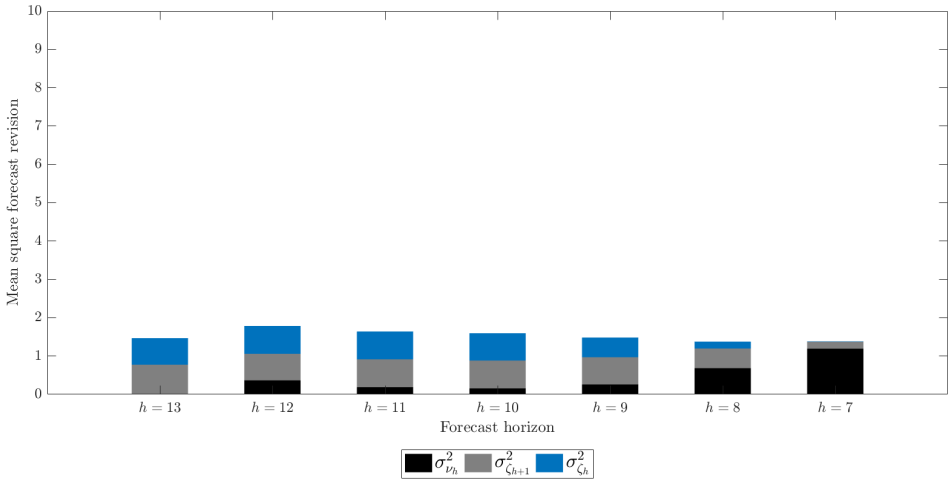
(b) Winter, before 22 May, 2012

Figure 3.10: MSFR and forecasting error components for summer and winter, before the NWP model upgrade.

Notes to Figure 3.10: The multi-horizon forecasts of maximum daily temperature produced by Stern and Davidson (2015) are best characterised by the rational-implicit model ($\hat{y}^{\beta}_{t|t-h} = \tilde{y}_t + \nu_{t|t-h} + \zeta_{t|t-h}$). See Table B.1 for parameter estimates, their standard errors, and AIC values for alternative models. Parameter $\sigma_{\omega_{h|t-h}}$ is the standard deviation of the rational forecast revision between horizon $h+1$ and horizon h. Parameter σ_{ζ_h} is the standard deviation of the implicit forecasting error component at horizon h. We interpret parameters σ_{ζ_h} as the inefficient use of information by forecasters. The information content of the multi-horizon forecasts, represented by σ_{ω}, is lower during the winter months before b) and after d) the Numerical Weather Prediction (NWP) model upgrade, than during the summer months before a) and after c) the NWP model upgrade. After the upgrade of the NWP model both summer and winter forecasts contain higher information content and higher implicit forecasting error.



(a) Summer, after 22 May, 2012



(b) Winter, after 22 May, 2012

Figure 3.11: MSFR and forecasting error components for summer and winter, after the NWP model upgrade.

Notes to Figure 3.11: The multi-horizon forecasts of maximum daily temperature produced by Stern and Davidson (2015) are best characterised by the rational-implicit model ($\hat{y}^{\beta}_{t|t-h} = \tilde{y}_t + \nu_{t|t-h} + \zeta_{t|t-h}$). See Table B.1 for parameter estimates, their standard errors, and AIC values for alternative models. Parameter $\sigma_{\omega_{h|t-h}}$ is the standard deviation of the rational forecast revision between horizon $h+1$ and horizon h. Parameter σ_{ζ_h} is the standard deviation of the implicit forecasting error component at horizon h. We interpret parameters σ_{ζ_h} as the inefficient use of information by forecasters. The information content of the multi-horizon forecasts, represented by σ_{ω}, is lower during the winter months before b) and after d) the Numerical Weather Prediction (NWP) model upgrade, than during the summer months before a) and after c) the NWP model upgrade. After the upgrade of the NWP model both summer and winter forecasts contain higher information content and higher implicit forecasting error.

3.6 Conclusion

Several approaches are available for evaluating the rationality of multi-horizon forecasts. This chapter develops a state-space approach for detecting a number of different types of forecast sub-optimality. Specifically, we decompose forecasting error into a rational component and a component related to mistakes introduced by the forecaster. We also extract a time varying estimate of forecast biasedness.

Both forecasters and forecast users may benefit from the insights gained by decomposing forecasting error in the way proposed by this chapter. The presence of implicit forecasting error or bias may alert forecasters to the presence of inefficiencies in their forecasting approach. Forecast users who recognise a trade-off between information content and information timeliness will benefit from an understanding of the marginal contribution of individual forecast revisions to the overall information content of the forecast series.

As an illustration, this chapter evaluates experimental multi-horizon forecasts of maximum daily temperature produced for Melbourne, Australia. We find that the forecasts are biased at all forecast horizons, with the amount of bias increasing as the forecast horizon increases. At short horizons the bias was found to be constant over time. The bias in longer horizon forecasts displayed cyclical behaviour, although the cyclical pattern was found to be irregular. We find evidence of forecast inefficiency at longer forecasts horizons. We find evidence that the upgrade of a NWP model used as an input to the forecasting system increased the information content of the forecast, while simultaneously increasing the size of forecasting mistakes.

The next chapter applies techniques developed in this chapter to evaluate multi-horizon demand forecasts in an electricity market.

Chapter 4

Understanding the information content of multi-horizon electricity demand forecasts

4.1 Introduction

Australia's National Electricity Market (NEM) is characterized by some of the most volatile wholesale spot prices in the world; see Weron (2014) for an overview. This huge network - the geographically longest in the world (Hurn et al., 2016), is facing threats to the security of domestic energy supply going forward. Increasing prices, infrastructure challenges, and environmental targets have raised sufficient concern to trigger a Governmental review of the future of the NEM due to report in 2017. The preliminary report released in December 2016 highlights the difficulties of forecasting consumer demand for electricity as one of the challenges to be faced (AEMC, 2015a).

The current wholesale electricity market for the NEM determines spot pricing through a market determined demand and supply spot market, with prices determined for every 5 minute interval. At 12:30pm each day generators provide

a bid of a segmented supply curve (with up to 10 separable price points) for the day commencing at 4:30am the following day. The Australian Electricity Market Operator matches the aggregated supply curve from the market with their estimate of market demand to produce a price signal. The estimated demand and price signals are fed back to the suppliers. Electricity markets differ from other commodity markets in that supply and demand must be instantaneously matched at each point in time, yet the commodity is non-storable. The majority of electricity generation capacity is provided by generators that have large fixed costs and low marginal costs. During periods of high demand additional capacity is met by generators that have increasingly higher marginal costs. Aggregating the supply curves of individual electricity generators reveals a market supply curve that is convex and steeply sloping; see for example the figures given in Hurn et al. (2016). Given the microeconomic fundamentals of this market, even a small unanticipated increase in demand, loss of supply, or both, can lead to a relatively large price increases. The frequency of large price increases in the NEM has led some market participants to suggest that electricity generators use their market power to increase price by strategically changing their offer to supply very close to when demand is realised. Several official inquires have produced only mixed evidence that generators possess market power and no evidence that generators engage in strategic behaviour by manipulating their offers to supply electricity (AEMC, 2013, 2015b).

However, there are some unique features of the Australian bidding process which theoretically lead to incentives for collusion. The first is that the 5 minute spot price is not the price realized by the wholesale generators. Instead they receive a half-hour average of these 5 minute prices. Second, the generators - who each bid supply - are able to alter the quantities (but not

prices) or their supply bids up to the 5 minute period prior to dispatch. This feature is unique to Australian markets. The only stipulation with regards to changes in supply is a good faith provision whereby initial offers and any subsequent changes must represent a genuine willingness to supply electricity at the nominated quantity and price. [1]

When the bids submitted by generators are provided in good faith they should represent the best available information about the supply conditions during each trading interval. Similarly, when AEMO's demand forecasts are efficient they should represent the best available information about the demand conditions during each trading interval. By matching the best available supply and demand information, price expectations during pre-dispatch should represent the best available information about price during each trading interval. The price expectations published by AEMO as part of the pre-dispatch process have been shown to contain significant bias and frequently underestimated actual spot price outcomes (Zainudin et al., 2015). Zainudin et al. (2015) conclude that the bias evident in pre-dispatch prices may be evidence of strategic supply side behaviour by the electricity generators.

This chapter investigates whether demand side information contained in the pre-dispatch process can explain these biased price expectations by evaluating the demand forecasts produced by AEMO for each trading interval. Drawing on the framework introduced in Chapter 3 of this thesis we apply the state space framework for forecast revision evaluation to the AEMO demand forecasts available to the markets. We find some evidence that the demand forecasts published by AEMO are biased, as well as significant evidence of

[1] From 2016 'late rebids' conducted in the 15 minutes prior to dispatch are required to retain documentation to rationalise their reasons for changes in supply. This process has not yet been tested.

forecast inefficiency. We present stylised facts on the information content of electricity demand forecasts and suggest that it is the pre-dispatch period, rather than the length of the forecast horizon, that plays the largest role in determining forecast information content.

The importance of this finding is that it points to the need to address both sides of the market in determining where better market outcomes can be achieved. While there is no doubt that there is considerable evidence that theoretically there may be incentives for suppliers in this market to behave strategically, there is, to our knowledge, no previous investigation of the means by which short-term revisions in demand conditions are incorporated into the market. This paper shows that there are demonstrable challenges to the market from the inefficiencies in demand-side forecasting.

4.2 Data

4.2.1 Structure of the pre-dispatch process

Figure 4.1 illustrates the relationship between calendar days, NEM trading days, trading intervals, and the pre-dispatch process. The electricity trading day begins at 4:01am each calendar day and runs until 4:00am the following calendar day. The trading day is divided into 48 half-hourly trading intervals with each trading interval further divided into 6 dispatch intervals of 5 minutes duration. We define electricity demand $D_{t,\tau}$ where t indexes the trading day and $\tau = 0430$ to 0400 indexes the end of each half hourly trading interval. The black line in Figure 4.1 represents the trading day and the 48 half hourly trading intervals. Demand $D_{t,2100}$ for 9:00pm on day t has been identified for

illustrative purposes. At the commencement of each dispatch interval AEMO observes demand and calls on generators along the aggregate supply curve, up until the point where demand and supply intersect, to dispatch the amount of electricity in their bid. The price during each dispatch interval is determined by the marginal generator's bid. AEMO then determines the spot price paid to generators by calculating the average price over the 6 dispatch intervals during the trading interval.

The pre-dispatch process for each trading day occurs at 12:30pm on the day before trading. Updated information from AEMO is relayed to the market every half hour throughout the trading day. For each of the pre-dispatch periods AEMO produces electricity demand forecasts for the half-hourly trading intervals.

To capture the pre-dispatch structure we define electricity demand forecasts produced on the day before trading day t as $\hat{D}_{t,\tau|t-1,j}$ where t indexes the trading day, $\tau = 0430$ to 0400 indexes the end of each half hourly trading interval, and j indexes the time of the pre-dispatch process. The blue lines in Figure 4.1 represent the 79 pre-dispatch processes for each trading day. There is an overlap between the time the pre-dispatch process for trading day t begins and the end of the pre-dispatch process for trading day $t-1$. We find it convenient to deal with these pre-dispatch process separately, and our notation makes this clear.

Demand forecast $\hat{D}_{t,\tau|t-1,2100}$, produced at 9:00pm on day $t-1$ has been identified for illustrative purposes. AEMO continues to revise their demand forecasts every half hour throughout the trading day for each of the remaining half-hourly trading intervals. We define electricity demand forecasts produced

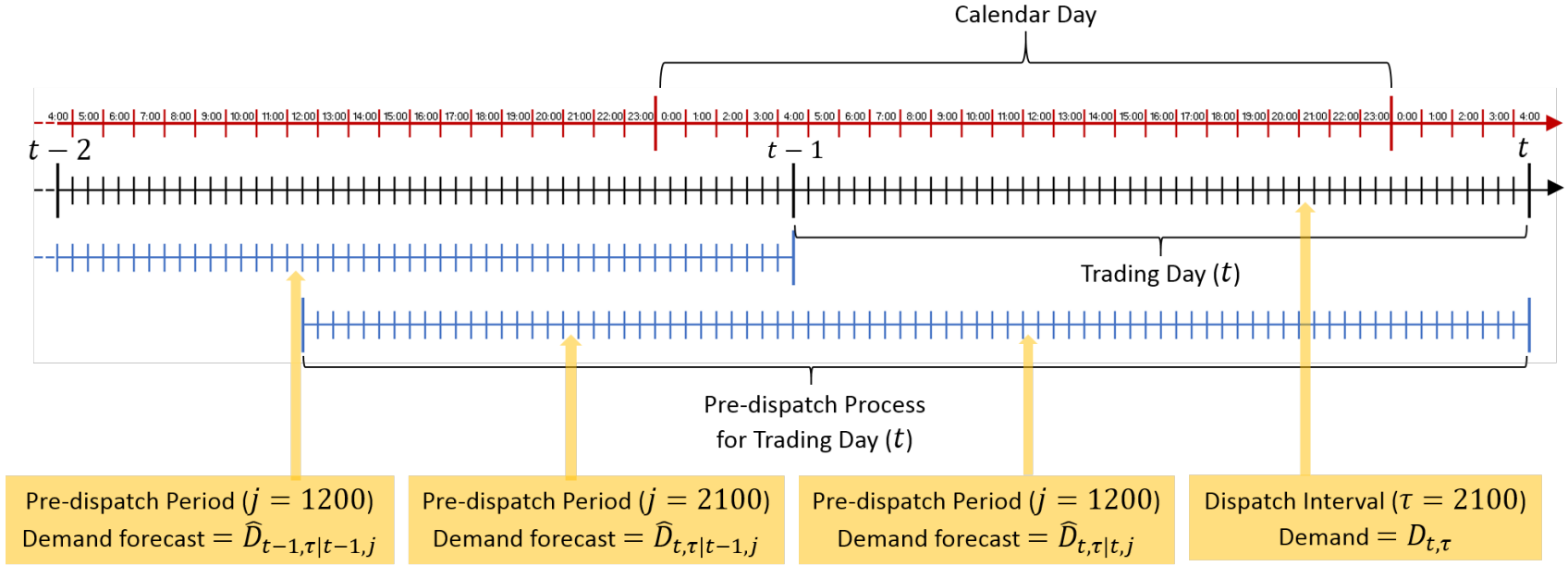


Figure 4.1: Schematic representation of the timing of electricity demand forecasts produced as part of AEMO's pre-dispatch process.

Notes to Figure 4.1: The pre-dispatch process for each trading day occurs at 12:30pm on the day before trading. Updated information from AEMO is relayed to the market every half hour throughout the trading day. For each of the pre-dispatch periods AEMO produces electricity demand forecasts for the half-hourly trading intervals. To capture this structure we define electricity demand forecasts produced on the day before trading day t as $\hat{D}_{t,\tau|t-1,j}$ where t indexes the trading day, $\tau = 0430$ to 0400 indexes the end of each half hourly trading interval, and j indexes the time of the pre-dispatch process. The blue lines represent the 79 pre-dispatch processes for each trading day. There is an overlap between the time the pre-dispatch process for trading day t begins and the end of the pre-dispatch process for trading day $t-1$. We find it convenient to deal with these pre-dispatch process separately, and our notation makes this clear. AEMO continues to revise their demand forecasts every half hour throughout the trading day for each of the remaining half-hourly trading intervals. We define electricity demand forecasts produced during trading day t as $\hat{D}_{t,\tau|t,j}$ where the indexing is the same as defined above.

during trading day t as $\hat{D}_{t,\tau|t,j}$ where the indexing is the same as defined above. The demand forecast $\hat{D}_{t-1,\tau|t-1,1200}$, produced at 12:00pm on day $t-1$, and demand forecast $\hat{D}_{t,\tau|t,1200}$, produced at 12:00pm on day t, have been identified for illustrative purposes.

As the pre-dispatch process commences on the day before trading, and runs throughout the trading day, demand forecasts for dispatch periods later in the trading day undergo more revisions than demand forecasts for trading periods earlier in the trading day. For instance, forecasts of $D_{t,0430}$, demand during the first trading interval of the trading day, are revised 31 times between the release of $\hat{D}_{t,0430|t-1,1230}$, at the time of the initial pre-dispatch process, and the release of $\hat{D}_{t,0430|t-1,0400}$ at 04:00am, just prior to the commencement of the trading day. Forecasts of $D_{t,0400}$, demand during the final trading interval of the trading day, are revised 78 times following the release of $\hat{D}_{t,0400|t-1,1230}$, at the time of the initial pre-dispatch process, and the release of $\hat{D}_{t,0400|t,0330}$ at 03:30am, just prior to the final trading interval of the trading day.

4.2.2 Forecast performance

We obtained from AEMO demand data for all half-hourly trading intervals from July, 2011 to July 2015 and demand forecasts produced as part of the corresponding pre-dispatch processes. This section uses mean absolute percentage error (MAPE) values to compare the performance of demand forecasts for all dispatch periods, produced during all of the pre-dispatch periods. MAPE is calculated as

$$100 \times |(\hat{D}_{t,\tau|t,j} - D_{t,\tau})|/D_{t,\tau}. \tag{4.1}$$

Heat maps in Figure 4.2 to Figure 4.8 present the MAPE values for these forecasts, and suggest that performance is related to both the time at which the forecasts are produced and the horizon at which the forecasts are generated. MAPE values show a similar intra-daily patterns across the week[2]. The weekdays have lower MAPE values, indicating higher forecast performance; possibly because electricity use patterns are more consistent during these periods. The highest intra-daily MAPE values, indicating lower forecast performance, occur from 12:00pm until around 7:00pm. On Wednesday and Friday higher MAPE values persist until later in the evening.

MAPE is likely to be influenced by the dispatch period being forecast, as well as the pre-dispatch period in which the forecast is produced. We expect MAPE to be higher at longer forecast horizons and to decline as the forecast horizon becomes shorter. This is property is implied by forecast rationality and has an intuitive appeal as we expect forecasts to contain more information at shorter forecast horizons.

Focusing on the forecasts for the Monday trading day in Figure 4.2, a movement from a long to a short forecast horizon is represented as a movement from left to right across the rows. The top row shows the 32 forecasts that are produced for the first dispatch interval of the trading day. The bottom row shows the 79 forecasts that are produced for the last dispatch interval of the trading day. The horizontal pattern of MAPE values from left to right across Figure 4.2 to Figure 4.8 broadly indicate that MAPE declines as the forecast horizon becomes shorter.

[2]Relatively high MAPE(%) values are observed during the first trading interval of Friday and during the middle of the day on Saturday and Sunday. We have confirmed that these apparent anomalies are a characteristic of the raw data, and not based on an error made during data cleaning or MAPE calculations. We leave further interpretation of this result for future work.

The vertical pattern of colour changes in Figure 4.2 to Figure 4.8 indicate that forecasts produced during the same pre-dispatch process have a similar decline in MAPE. One explanation for why the time of pre-dispatch matters for forecast performance could be that information available to forecasters is lumpy. That is, demand forecasts are produced half-hourly to update the pre-dispatch process regardless of whether new information is available. From a practical perspective information relevant to market conditions over several trading intervals may only arrive at specific periods throughout the day. For example, weather forecasts, used as predictors in electricity load forecasting models, may be updated by meteorologists at a lower frequency than the half hourly pre-dispatch forecast revisions.

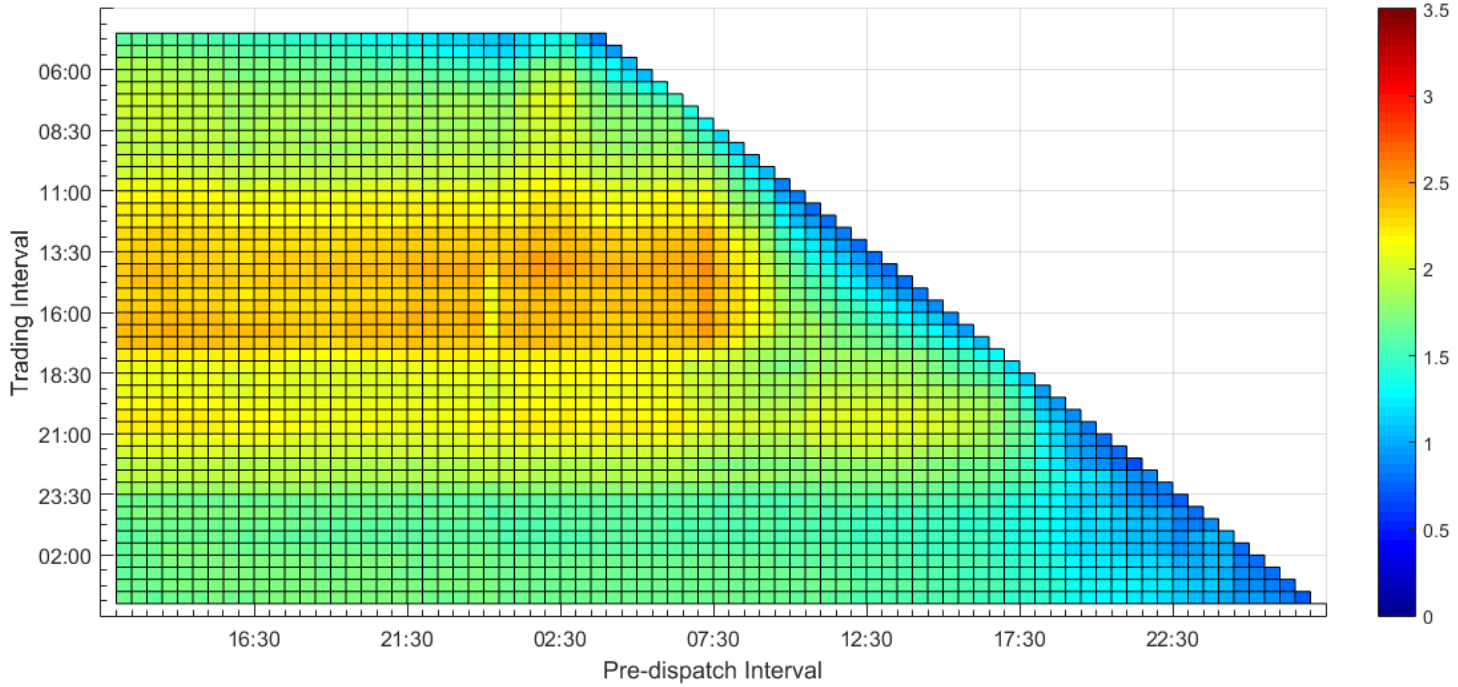


Figure 4.2: Monday electricity trading day — MAPE(%) heat map of demand forecast performance.

Notes to Figure 4.2: Data are from the Australian Energy Market Operator (AEMO). AEMO produce half-hourly electricity demand forecasts as part of the pre-dispatch process commencing at 12:30pm on the day prior to trading. The forecasts being evaluated are for the News South Wales region of the National Electricity Market (NEM), over the period July 3, 2011, to July 31, 2015. Mean Absolute Percentage Error (MAPE) is calculated as $100 \times |(\hat{D}_{t,\tau|t,j} - D_{t,\tau})|/D_{t,\tau}$ for forecasts produced during each of the half-hourly pre-dispatch periods.

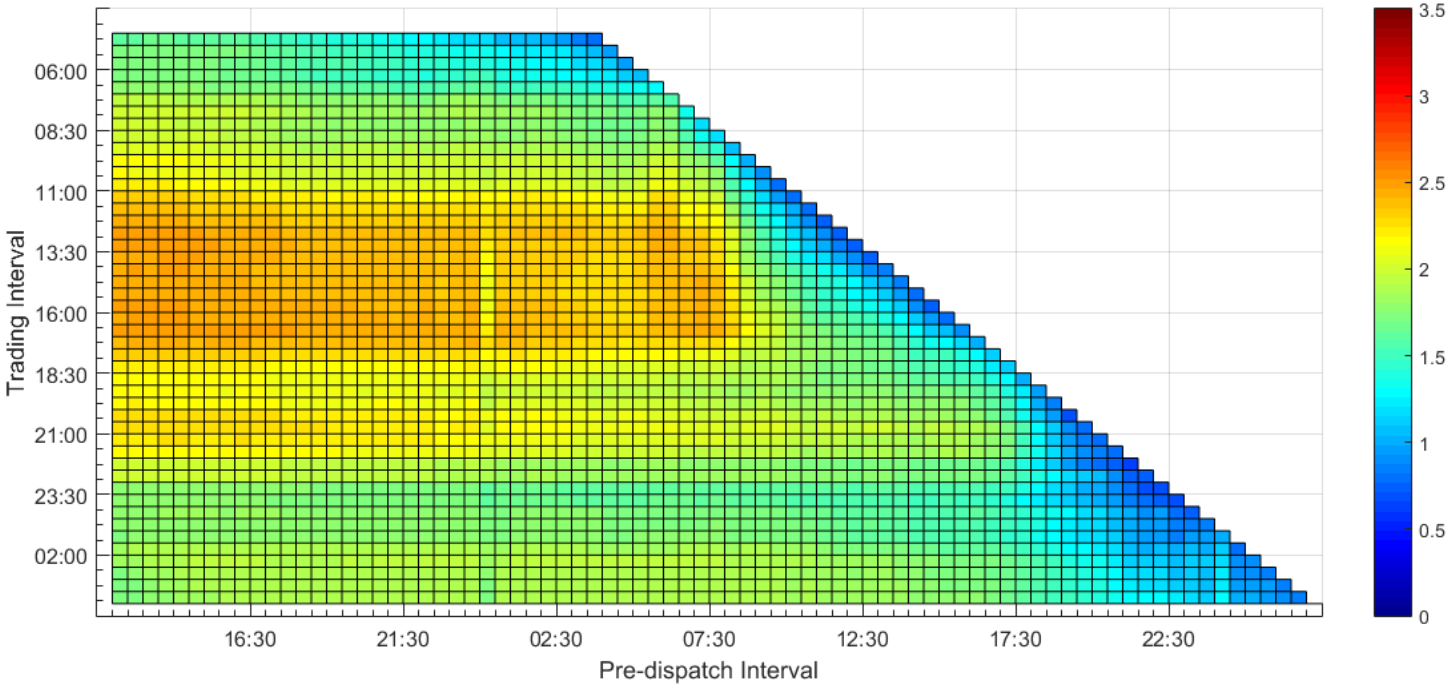


Figure 4.3: Tuesday electricity trading day — MAPE(%) heat map of demand forecast performance.

Notes to Figure 4.3: see notes to Figure 4.2.

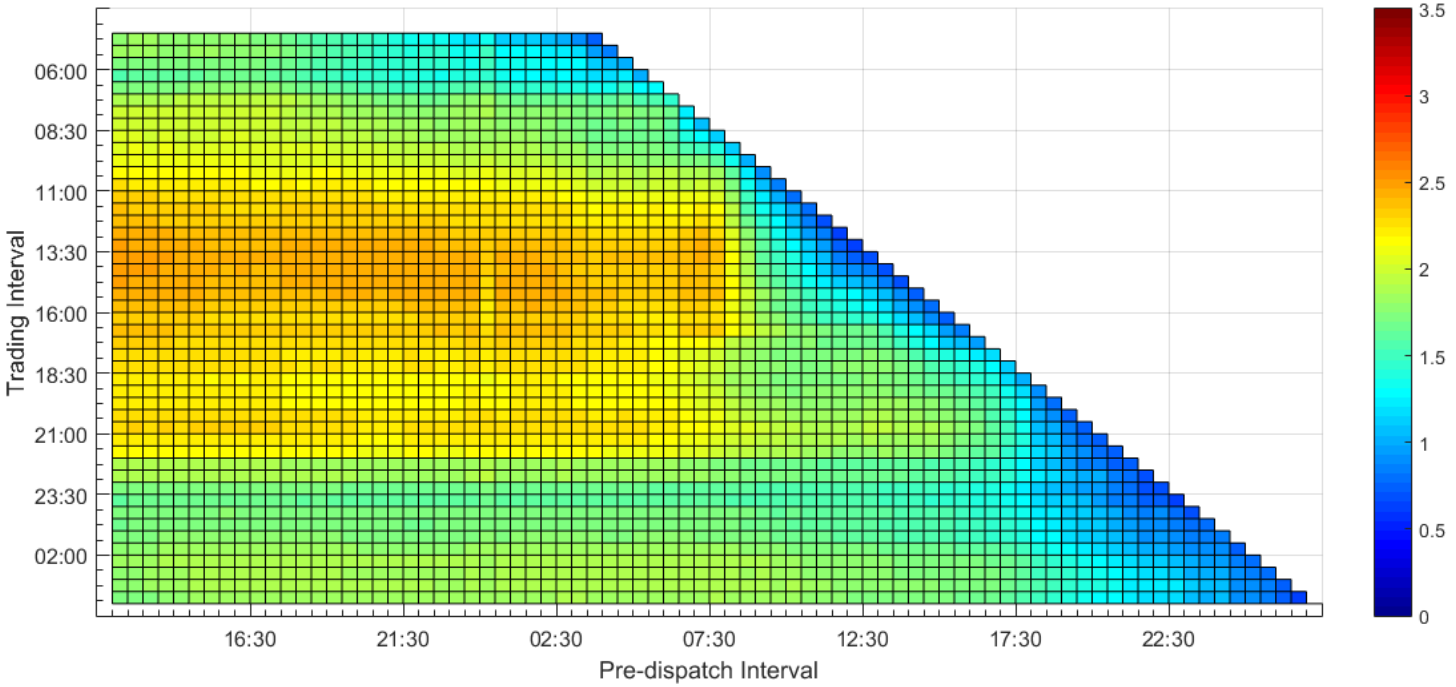


Figure 4.4: Wednesday electricity trading day — MAPE(%) heat map of demand forecast performance.

Notes to Figure 4.4: see notes to Figure 4.2.

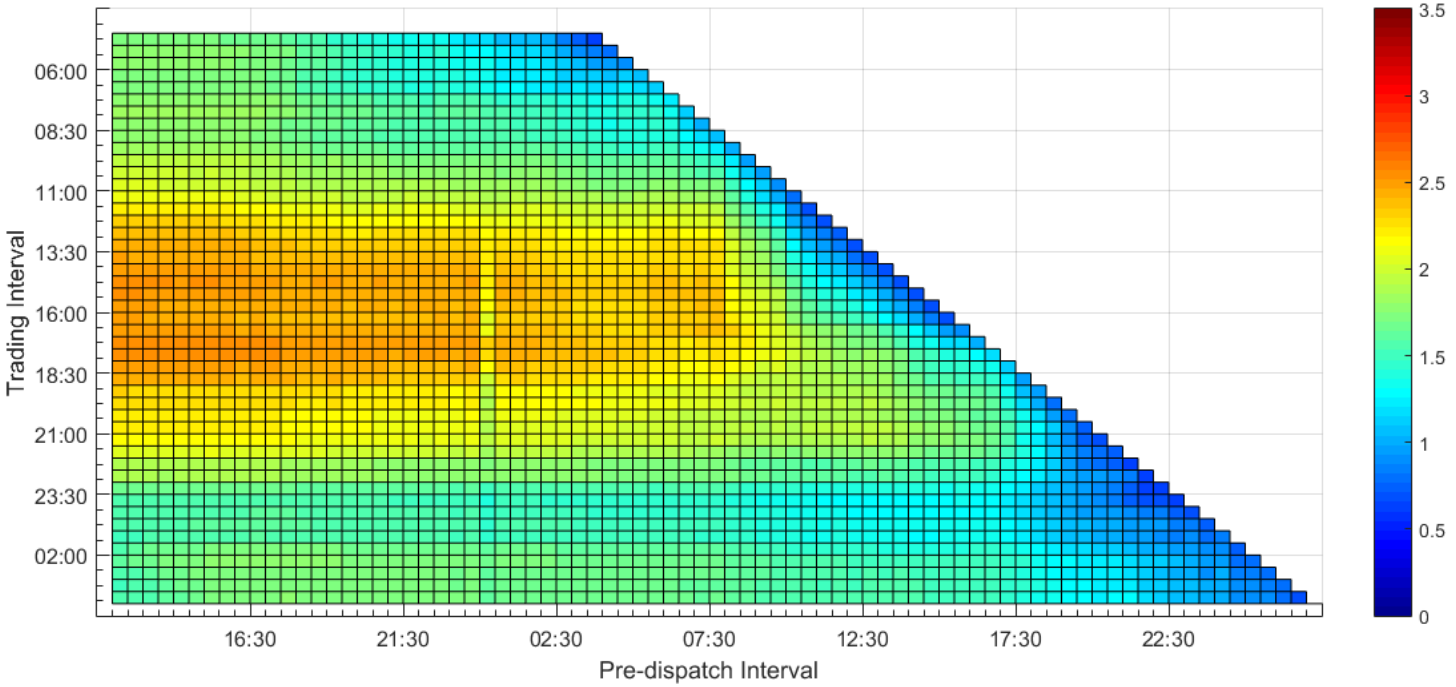


Figure 4.5: Thursday electricity trading day — MAPE(%) heat map of demand forecast performance.

Notes to Figure 4.5: see notes to Figure 4.2.

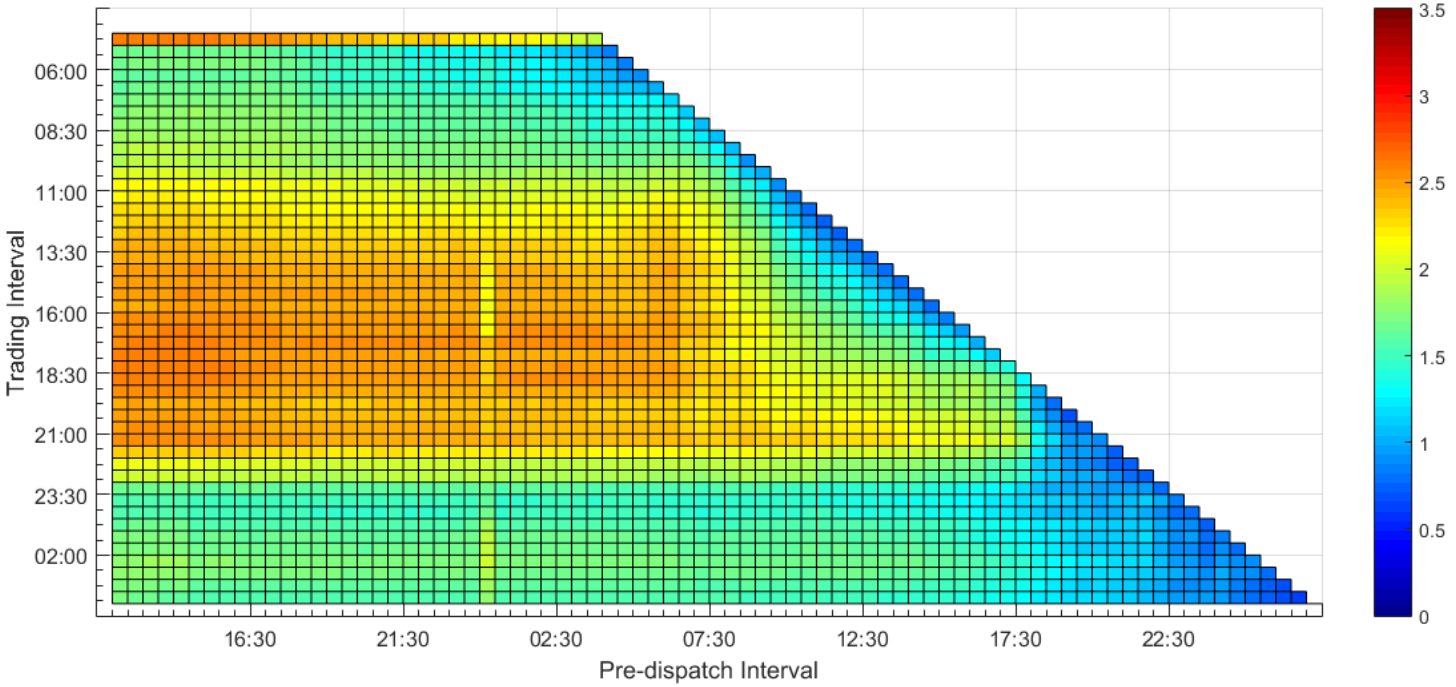


Figure 4.6: Friday electricity trading day — MAPE(%) heat map of demand forecast performance.

Notes to Figure 4.6: see notes to Figure 4.2.

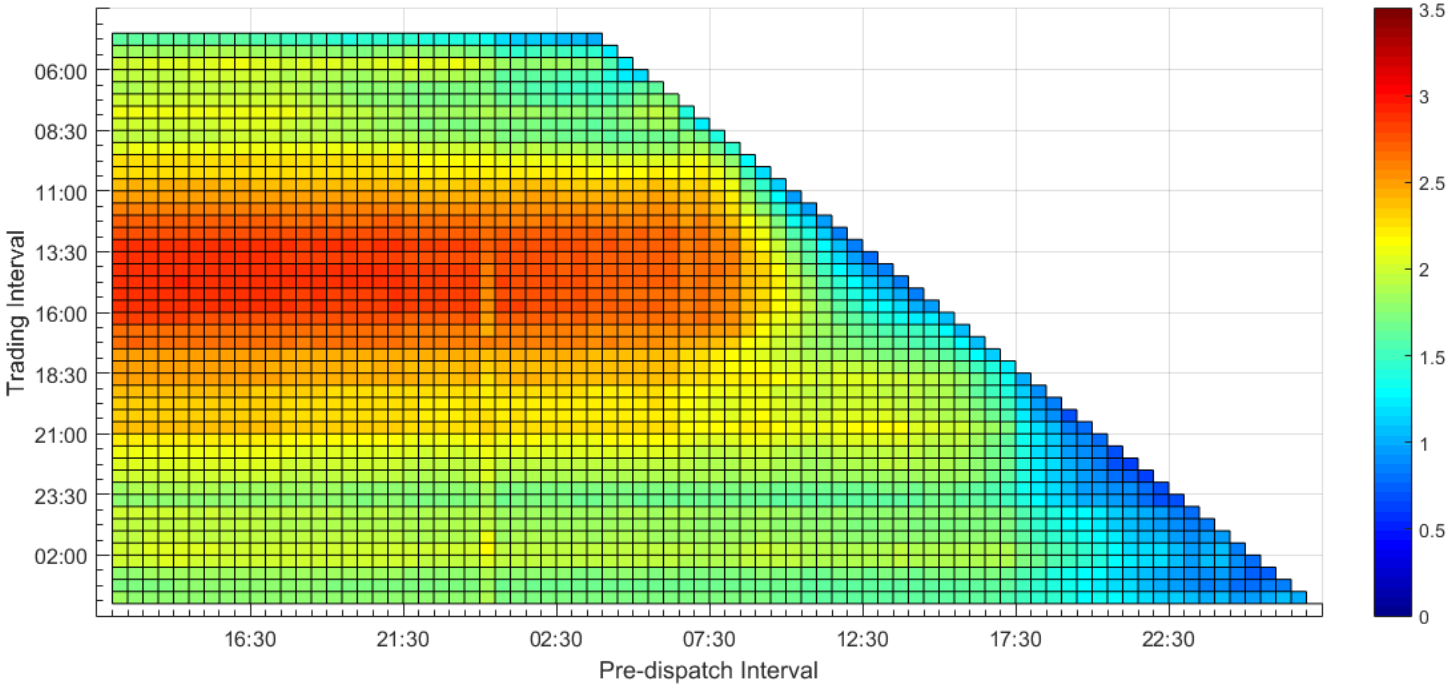


Figure 4.7: Saturday electricity trading day — MAPE(%) heat map of demand forecast performance.

Notes to Figure 4.7: see notes to Figure 4.2.

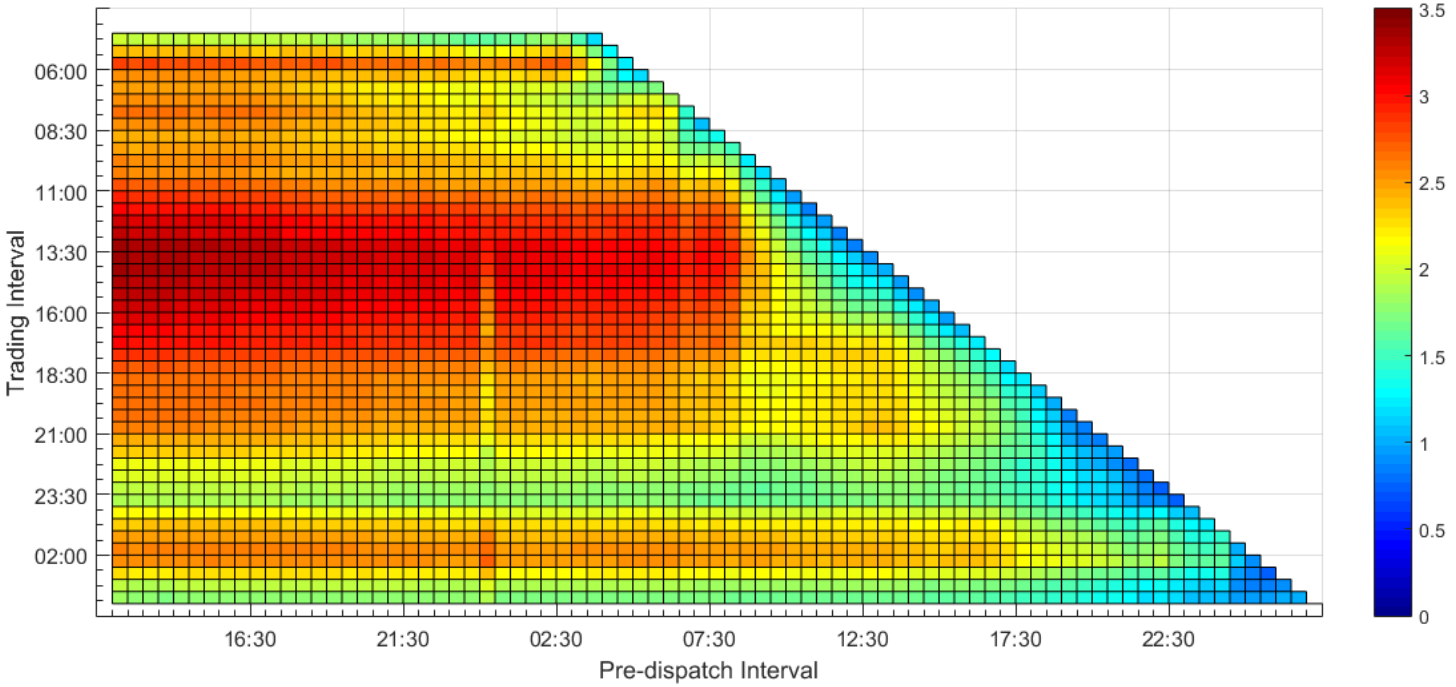


Figure 4.8: Sunday electricity trading day — MAPE(%) heat map of demand forecast performance.

Notes to Figure 4.8: see notes to Figure 4.2.

4.3 Method

We evaluate AEMO's electricity demand forecasts using the state-space appraoch to multi-horizon forecast evaluation developed in Chapter 3.

First, we estimate the level of non-zero trend in the pre-dispatch electricity demand forecasting error using a process analogous to equations 3.32 to 3.34, where forecasting error is calculated as

$$e_{t,\tau|t,j} = D_{t,\tau} - \hat{D}_{t,\tau|t,j}. \tag{4.2}$$

Any non-zero trend in the pre-dispatch electricity demand forecasting error is then extracted from the forecast variables.[3].

Next, we use our model of multi-horizon forecasts to evaluate the unbiased pre-dispatch demand forecasts. Our most general model assumes that forecasts contain rational and implicit components

$$\hat{D}_{t,\tau|t,\tau-h} = \tilde{D}_{t,\tau} + \nu_{t,\tau|t,\tau-h} + \zeta_{t,\tau|t,\tau-h} \tag{4.3}$$

where the conditioning on $t, \tau - h$, for $h \geq 0$ corresponding to the half-hourly pre-dispatch intervals, indicates that the forecast components are horizon specific.

When equation (4.3) is cast in state space form the measurement equation describes the multi-horizon forecasts and the target variable

$$\boldsymbol{y}_t = [\hat{D}_{t,\tau|t,\tau-h}, \ldots, \hat{D}_{t,\tau|t,\tau-1}, D_{t,\tau}]' \tag{4.4}$$

in terms of a dynamic representation of the target variable, rational forecasting error, and implicit forecasting error, as follows

[3]The unbiased forecasts are defined by $\hat{D}_{t,\tau|t,j} - \hat{\beta}_{t,\tau|t,j}$, where $\hat{\beta}_{t,\tau|t,j}$ is the trend in $e_{t,\tau|t,j}$

$$\boldsymbol{y}_t = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{Z}_2 & \boldsymbol{Z}_3 \end{bmatrix} \cdot \begin{bmatrix} \tilde{D}_{t,\tau} \\ \boldsymbol{\nu}'_{t,\tau|t,\tau-h} \\ \boldsymbol{\zeta}'_{t,\tau|t,\tau-h} \end{bmatrix} \tag{4.5}$$

where $\tilde{D}_{t,\tau}$ is a dynamic representation of the target variable, $\boldsymbol{\nu}_{t,\tau|t,\tau-h}$ is horizon specific rational forecasting error, and $\boldsymbol{\zeta}_{t,\tau|t,\tau-h}$ is horizon specific implicit forecasting error. The matrices $\{\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3\}$ are $\mathbf{1}_{(h+1)}$, $\left[\begin{smallmatrix} \boldsymbol{I}_h \\ \mathbf{0} \end{smallmatrix}\right]$, and $\left[\begin{smallmatrix} \boldsymbol{I}_h \\ \mathbf{0} \end{smallmatrix}\right]$.

The state equation describes the dynamics of the forecast components, as follows

$$\begin{aligned} \begin{bmatrix} \tilde{D}_{t,\tau} \\ \boldsymbol{\nu}'_{t,\tau|t,\tau-h} \\ \boldsymbol{\zeta}'_{t,\tau|t,\tau-h} \end{bmatrix} &= \begin{bmatrix} T_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{T}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_3 \end{bmatrix} \cdot \begin{bmatrix} \tilde{D}_{t-1,\tau} \\ \boldsymbol{\nu}'_{t-1,\tau|t-1,\tau-h} \\ \boldsymbol{\zeta}'_{t-1,\tau|t-1,\tau-h} \end{bmatrix} \\ &+ \begin{bmatrix} R_1 & \boldsymbol{R}_2 & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{U} \cdot \mathrm{diag}(\boldsymbol{R}_2) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathrm{diag}(\boldsymbol{R}_3) \end{bmatrix} \cdot \begin{bmatrix} \eta_{\xi,t,\tau} \\ \boldsymbol{\eta}'_{\omega_h,t,\tau} \\ \boldsymbol{\eta}'_{\zeta_h,t,\tau} \end{bmatrix} \end{aligned} \tag{4.6}$$

where $T_1 = 1$, $\boldsymbol{T}_2 = \boldsymbol{T}_3 = \mathbf{0}_h$, and $\mathbf{0}_h$ are conformably defined null matrices. The selection matrix $\boldsymbol{R}$ describing the dynamics of the state equation disturbance terms where $R_1 = \sigma_\xi$ and $\boldsymbol{R}_2 = [\sigma_{\omega_h}, \sigma_{\omega_{h-1}}, \ldots \sigma_{\omega_1}]$ together form the standard deviation of the innovations associated with the dynamic representation of the target variable, $\boldsymbol{U}$ is an upper triangular matrix of ones, $\boldsymbol{R}_3 = [\sigma_{\zeta_h}, \sigma_{\zeta_{h-1}}, \ldots \sigma_{\zeta_1}]$ are standard deviations of the implicit innovations. Finally, $\boldsymbol{\eta} \sim$ i.i.d. $\mathrm{N}(0, 1)$.

4.4 Results

We focus on evaluating electricity demand forecasts produced during the last eight pre-dispatch processes prior to each of the 48 trading intervals during each of trading day of the week. This constrains the extent of the problem, although future work will expand this horizon. Evaluating eight forecast

horizons for each trading interval of each day of the week represents 2,688 individual forecasts per week. Examining such a large number of forecasts allows us to explore the influence of daily and weekly seasonality on forecast performance. Slicing the data in this way is likely to be important as the parameters of electricity demand forecasting models are often estimated separately for individual trading periods (Taieb and Hyndman, 2014; Black and Henson, 2014; Fan and Hyndman, 2012; Ramanathan et al., 1997).

We examine demand forecasts using the rational and rational-implicit modelling frameworks. Forecast rationality requires that forecasts variance be less than the variance of the target variable. This assumption underpins the rational component of our model. Prior to estimating the models we compared the variance of forecasts at each horizon to the variance of the target variable. The variance of AEMOs pre-dispatch electricity demand forecasts during most pre-dispatch intervals exceeds the variance of actual demand for almost all dispatch periods. This suggests that the rational model of multi-horizon forecasts would be a poor fit for the data. Our empirical investigations established that this was indeed the case. Consequently, the rational model of multi-horizon forecasts will not be discussed further and the remainder of the chapter deals with the rational-implicit model.

The results section proceeds as follows. In subsection 4.4.1 we present evidence that AEMO demand forecasts for some dispatch periods contain forecast bias. In subsection 4.4.2 we examine the information content of the pre-dispatch process using estimates of the rational forecasting error component.

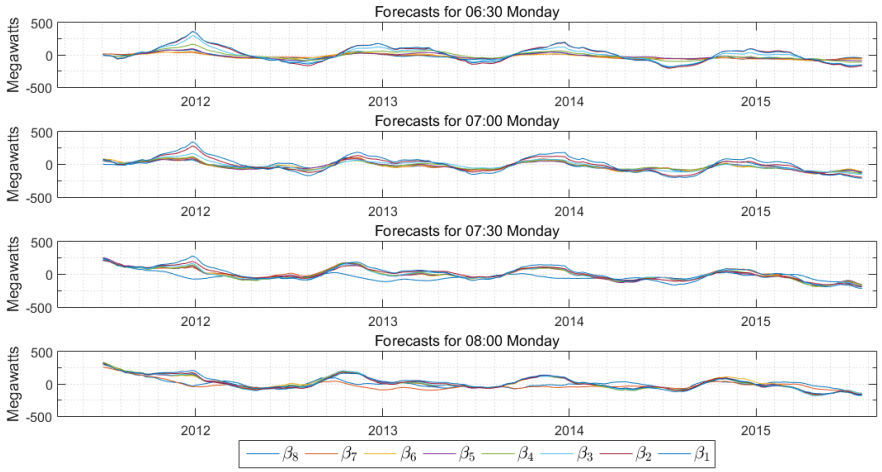
4.4.1 Forecast bias

Figure 4.9 presents two examples of bias estimates in forecasts for the Monday trading day: early morning (4:30am to 8:00am) and late evening (8:30pm to 10:00pm). Forecasts for these 16 trading intervals contain a higher amount of bias than other periods during the Monday trading day. A similar pattern is seen for bias estimates in the forecasts for the trading intervals during other days of the week (see Appendix D).

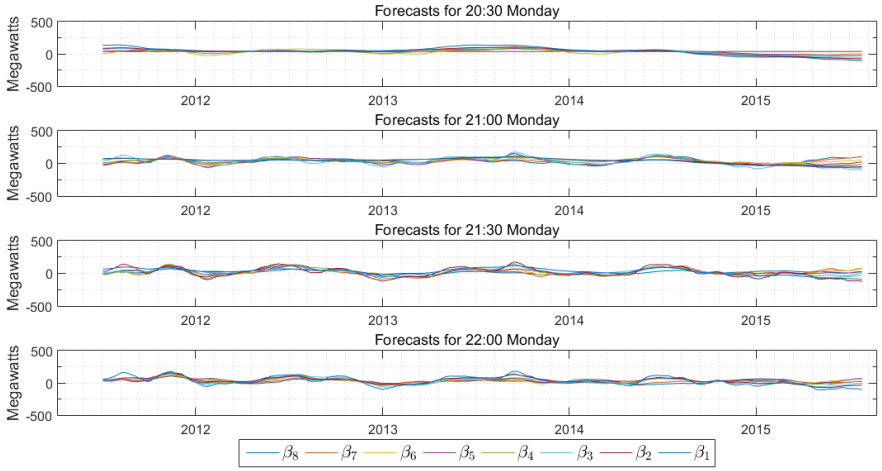
AEMO produce alternative demand forecast scenarios to evaluate how sensitive their price forecasts are to changes in demand based around scenarios of +/- 200 megawatts. Our results show that bias in the forecasts for several trading intervals between 4:30am and 12:00pm is estimated to be larger than 200 megawatts. While bias is largely absent in forecasts during the afternoon and early evening, bias is estimated to increase again after 9:00pm until the end of the trading day. There also appears to be a seasonal pattern in the bias estimates. Bias in the forecasts for the morning trading intervals is highest during the warmer months and largely absent during the cooler months. Bias in the forecasts for the late evening trading intervals is highest during the warmest and the coolest months of the year.

4.4.2 Information content

We estimate two models: rational; and rational-implicit. The variance of AEMOs pre-dispatch electricity demand forecasts during most pre-dispatch intervals exceeds the variance of actual demand for almost all dispatch periods. This suggests that the rational model of multi-horizon forecasts would be a



(a)



(b)

Figure 4.9: Two examples of bias estimates in electricity demand forecasts for the Monday trading day.

Notes to Figure 4.9: Data are from the Australian Energy Market Operator (AEMO). AEMO produce half-hourly electricity demand forecasts as part of the pre-dispatch process commencing at 12:30pm on the day prior to trading. The forecasts being evaluated are for the News South Wales region of the National Electricity Market (NEM), over the period July 3, 2011, to July 31, 2015. Bias as reported here is the state component estimate from a local linear trend model applied to forecasting error $e_{t,\tau|t,j} \equiv \hat{D}_{t,\tau|t,j} - D_{t,\tau}$, where $j = \tau - i$, for $i = 1, \ldots, 8$.

poor fit for the data. This was indeed the case.

Figures 4.10 to 4.14 show that the time of pre-dispatch plays a larger role in determining the information content of forecast revisions than the horizon at which the forecast is produced. The information content of forecast revisions, as recorded on the y-axis, is the estimated standard deviation of rational forecast revisions σ_{ω_i} from the rational-implicit model scaled by the standard deviation of the corresponding target variable series $\sqrt{\text{Var}(D_{t,\tau})}$. Scaling σ_{ω_i} in this way allows us to compare different τ trading intervals throughout the electricity trading day as well as different days of the week. We define this scaled variable as $\text{S}\sigma_{\omega_i}$. We expect the variability of the forecasts to increase as they get revised with new information. Recall that σ_{ω_i} represents the variability of the rational forecast revision between horizon $h = i$ and $h = i - 1$. Therefore the lines $\text{S}\sigma_{\omega_1}$ to $\text{S}\sigma_{\omega_7}$ represent the information content introduced during the pre-dispatch process.

The x-axis refers to the time of the pre-dispatch process, not the trading interval, and so indicates the time that the information content is realised. Prior to 07:00am the length of the forecast horizon appears to be the main factor determining the information content of the forecast revisions. This is evident from the spread of the lines representing the different forecast horizons. At 07:00am the lines for all forecast revisions except $\text{S}\sigma_{\nu_7}$ bunch together until around midday. This suggests that the information content of the forecast revisions during the morning pre-dispatch periods are similar for all trading intervals, regardless of the length of the forecast horizon. After midday the lines separate again suggesting the forecast horizon once again becomes the main factor determining the information content of forecasts. The $\text{S}\sigma_{\omega_i}$ lines once again bunch together during late afternoon and into the evening as the

information content is again determined by the pre-dispatch period in which the revisions are occurring.

While there are some differences in the amount of information content across the days of the week the pattern of information content throughout the day appears broadly similar.

4.4.3 Market design — Policy Implications

The results of this investigation into the demand side forecasting component of the electricity market strongly suggest that there are gains to be made here as well as those posited on the supply side of the market. Inefficient demand forecasts introduce error into the pre-dispatch price expectations, and will thereby contribute to the problems in delivering reliable, well-priced supply smoothly. The presence of biased demand forecasts provide an alternative, not necessarily mutually exclusive, explanation for the presence of biased pre-dispatch price expectations. As far as we are aware this is the first examination of the performance of the pre-dispatch demand-side forecasting for the purposes of price determination, and the results so far indicate that it is worthy of further investigation in refining the market operation of the NEM in Australia.

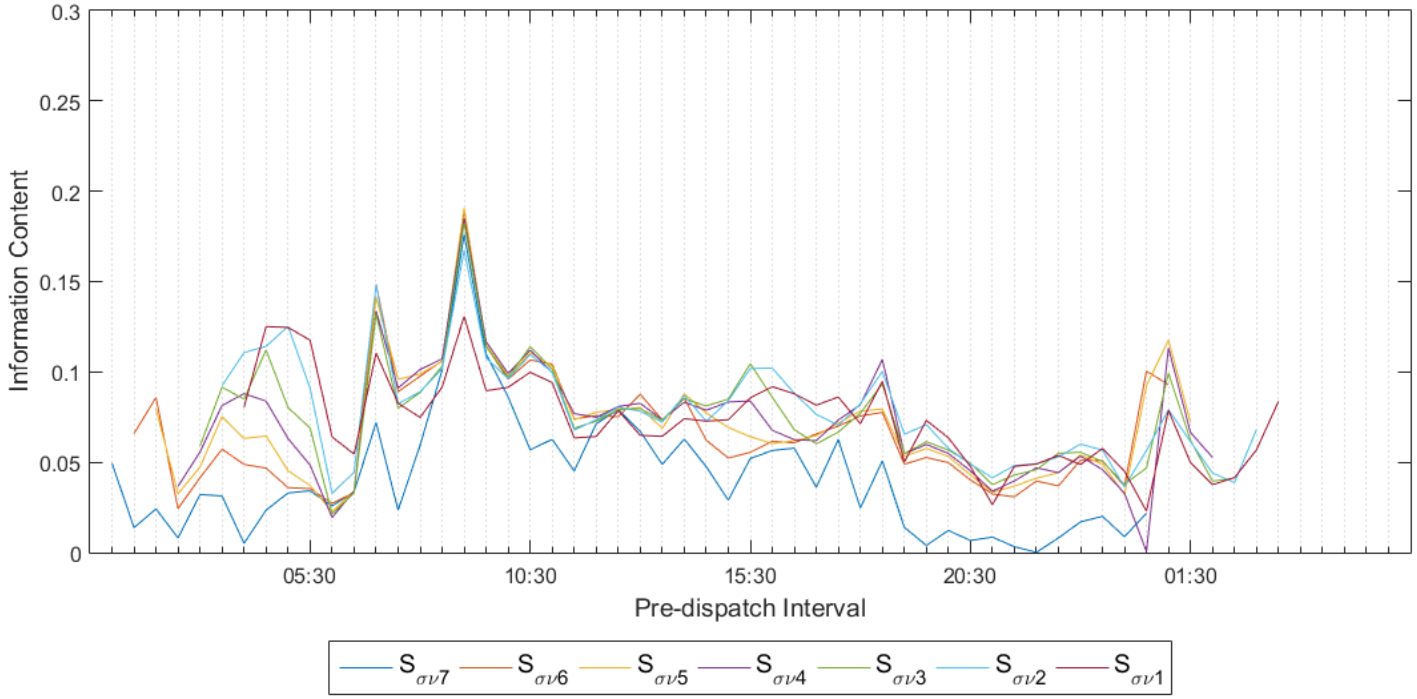


Figure 4.10: Monday electricity trading day — information content demand forecast revisions.

Notes to Figure 4.10: Data are from the Australian Energy Market Operator (AEMO). AEMO produce half-hourly electricity demand forecasts as part of the pre-dispatch process commencing at 12:30pm on the day prior to trading. The forecasts being evaluated are for the News South Wales region of the National Electricity Market (NEM), over the period July 3, 2011, to July 31, 2015. The information content of forecast revisions, as recorded on the y-axis, is the estimated standard deviation of rational forecast revisions σ_{ω_i} from the rational-implicit model scaled by the standard deviation of the corresponding target variable series $\sqrt{\text{Var}(D_{t,\tau})}$. Scaling σ_{ω_i} in this way allows us to compare different τ trading intervals throughout the electricity trading day as well as different days of the week. The x-axis refers to the time of the pre-dispatch process, not the trading interval, and so indicates the time that the information content is realised.

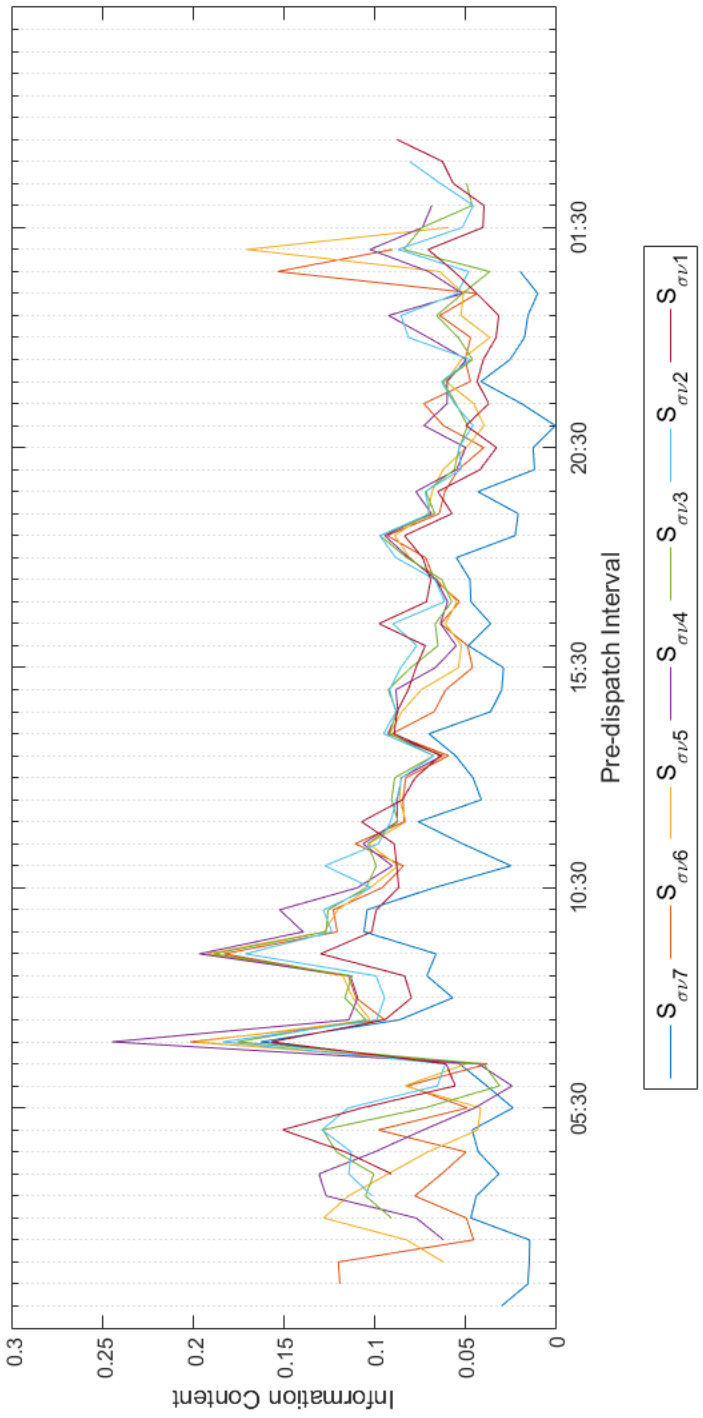


Figure 4.11: Tuesday electricity trading day — information content demand forecast revisions.

Notes to Figure 4.11: see notes to Figure 4.10.

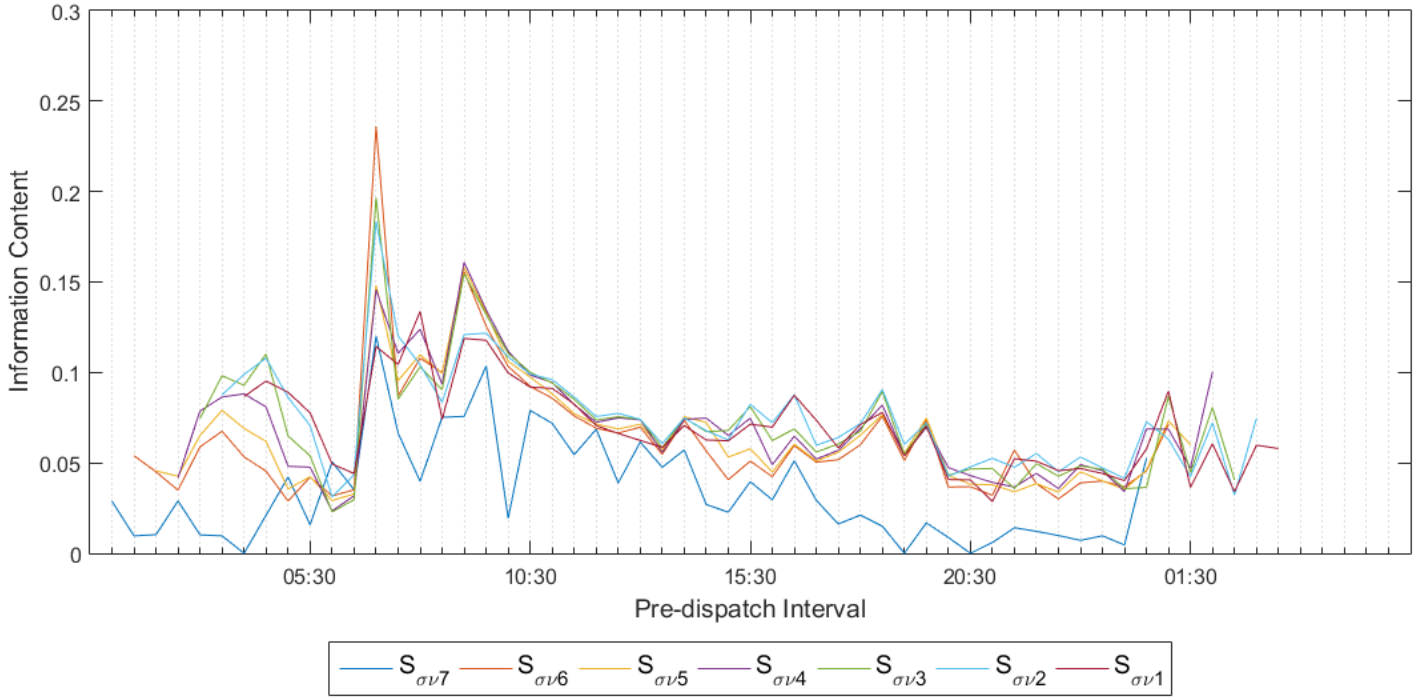


Figure 4.12: Wednesday electricity trading day — information content demand forecast revisions.

Notes to Figure 4.12: see notes to Figure 4.10.

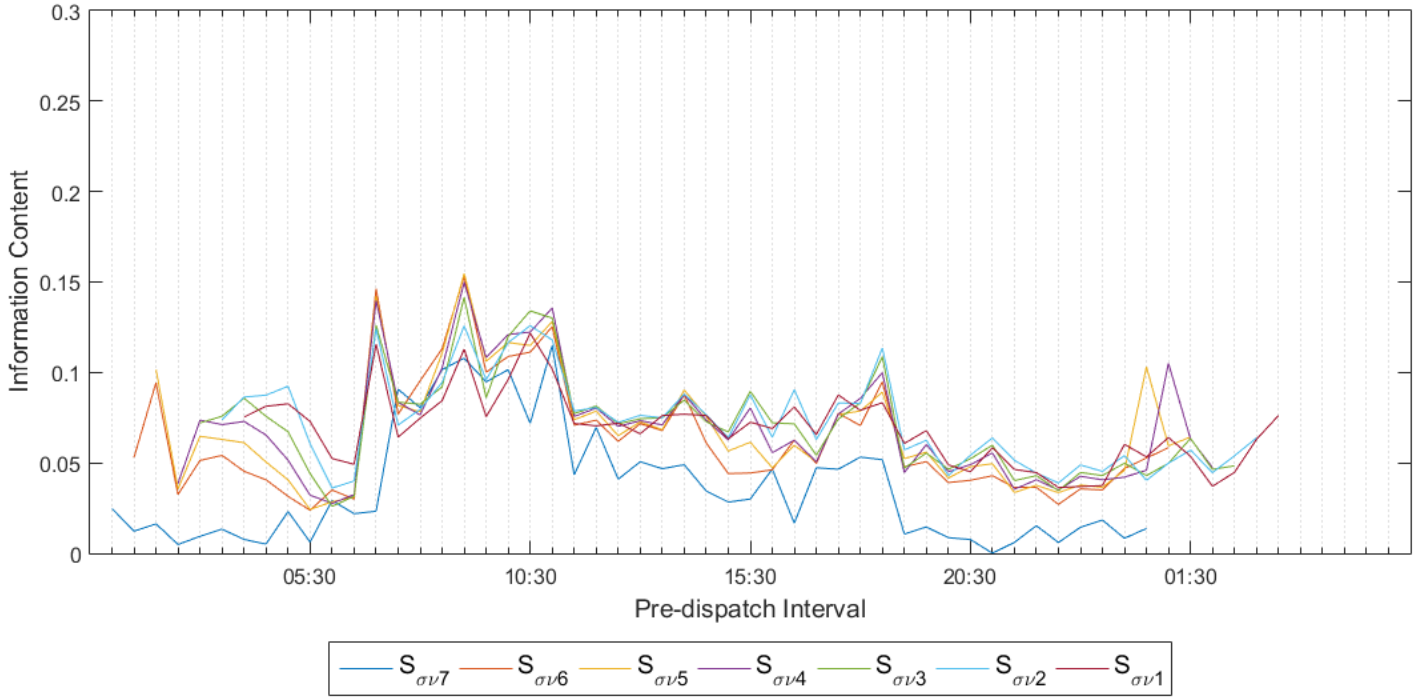


Figure 4.13: Thursday electricity trading day — information content demand forecast revisions.

Notes to Figure 4.13: see notes to Figure 4.10.

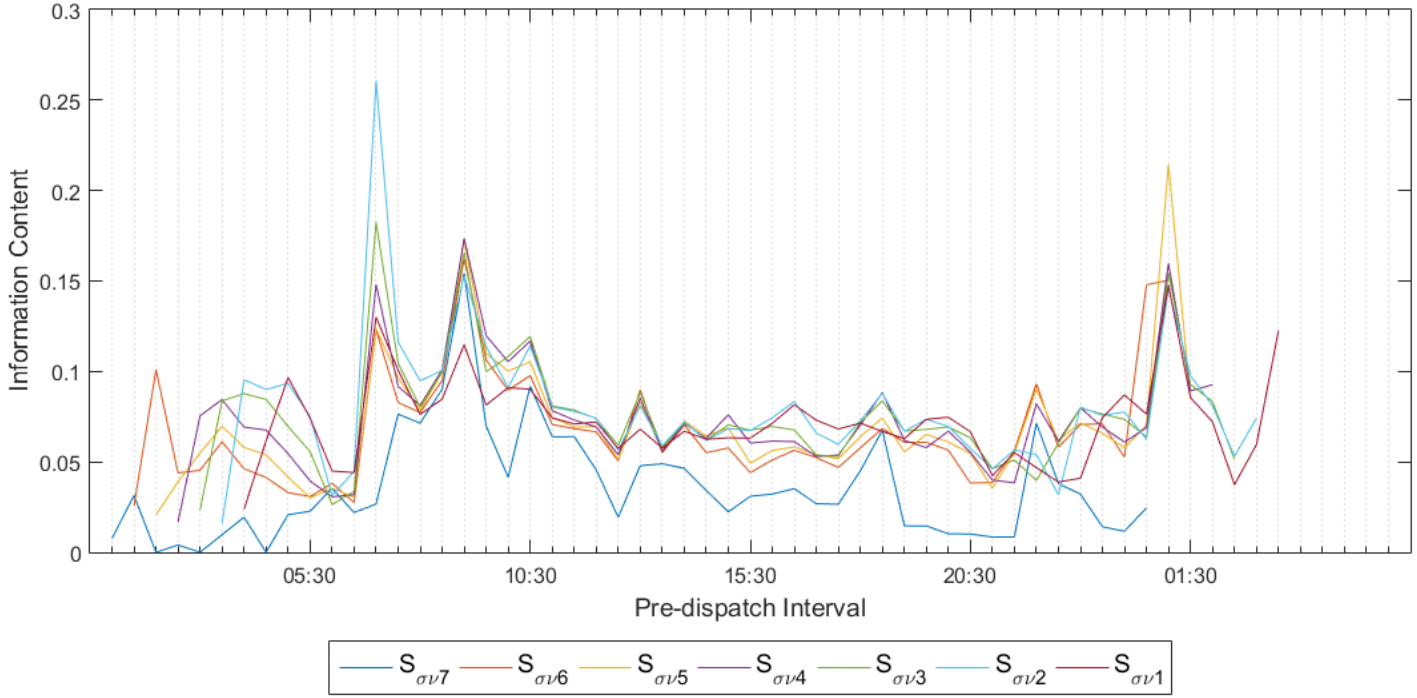


Figure 4.14: Friday electricity trading day — information content demand forecast revisions.

Notes to Figure 4.14: see notes to Figure 4.10.

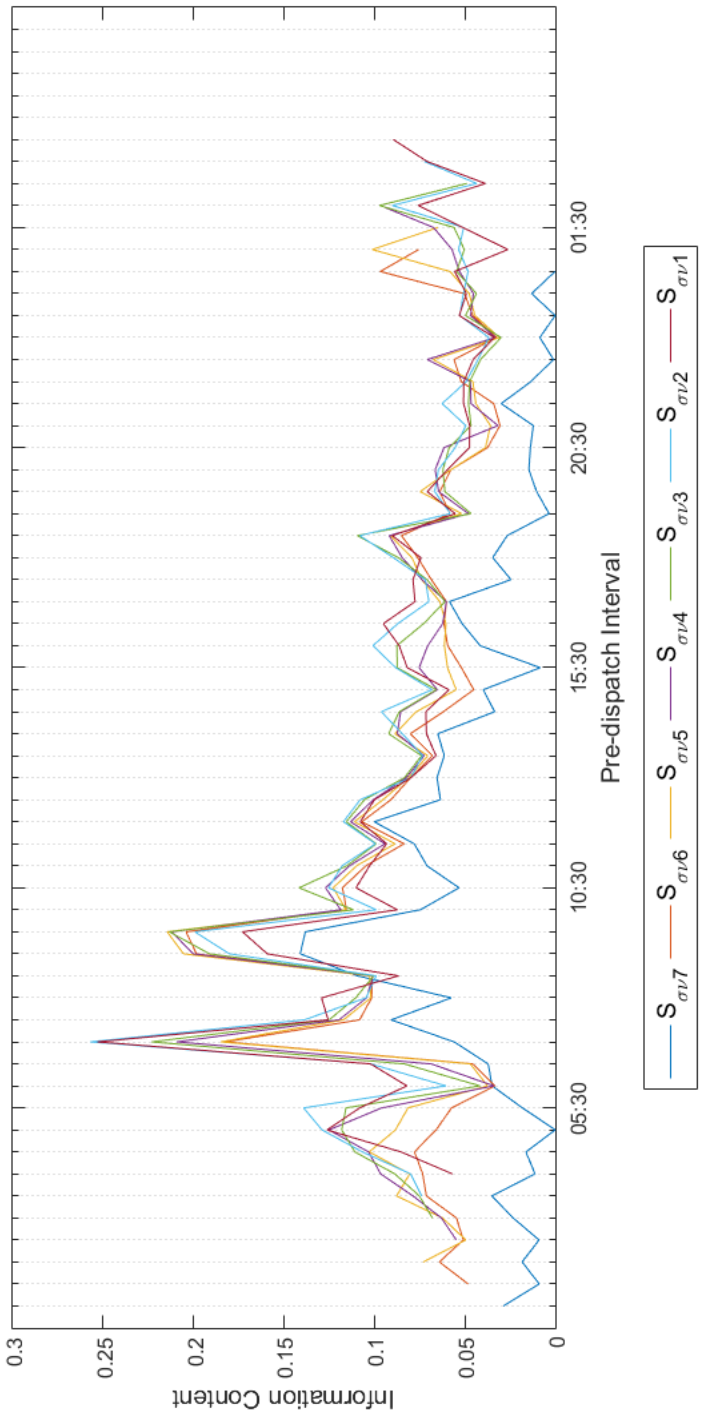


Figure 4.15: Saturday electricity trading day — information content demand forecast revisions.

Notes to Figure 4.15: see notes to Figure 4.10.

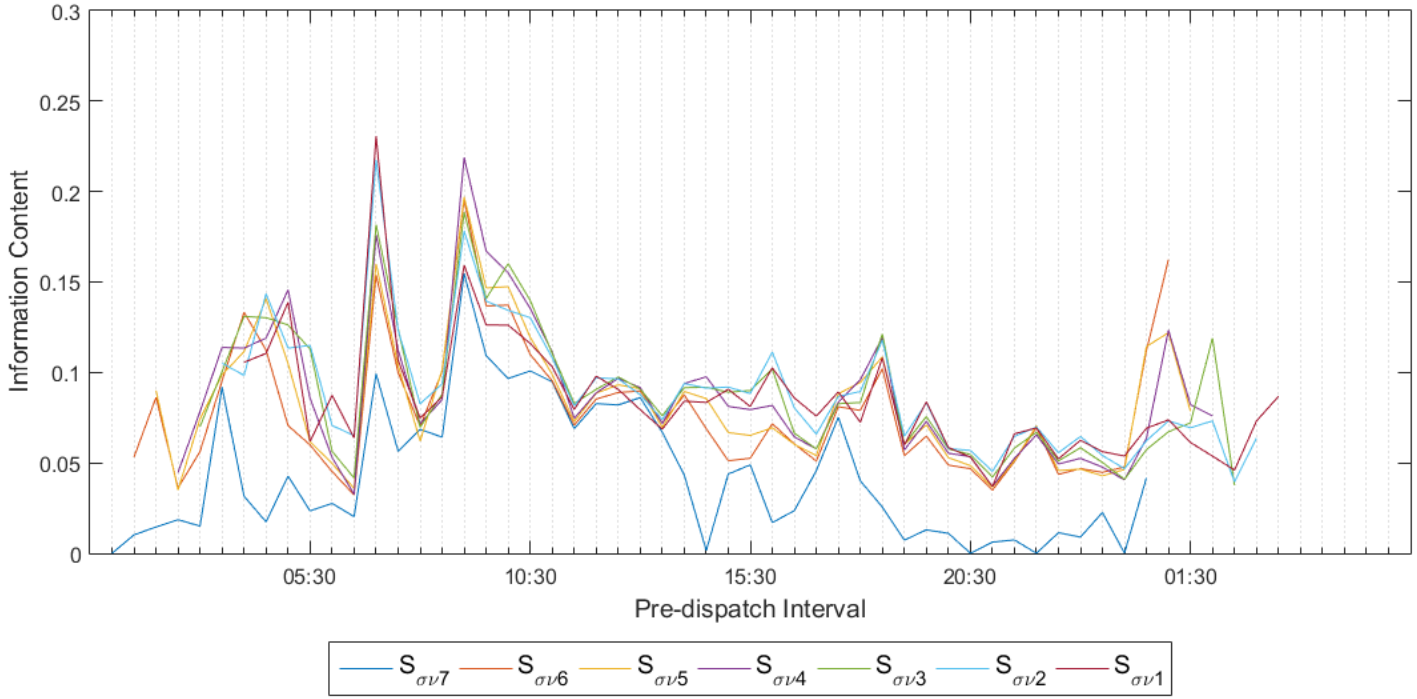


Figure 4.16: Sunday electricity trading day — information content demand forecast revisions.

Notes to Figure 4.16: see notes to Figure 4.10.

4.5 Conclusion

This chapter provides an illustration of the real-world applications where the forecast revision assessment methodology developed in Chapter 3 of this thesis can be usefully implemented to extract information about the sources of updated information, and in this case how this may impact on important market outcomes. The wholesale price of electricity is of great importance to the successful operation of the economy, both for producers and households, all of whom consume electricity, and to governments tasked with maintaining security and reliability of supply, as well as meeting potentially competiting environmental and investment aims. In future work we aim to consider how the information from the demand and supply side can take advantage of the evolution of information in the revisions from both sides to improve outcomes in this market. The aim is to contribute to the the current enquiries on means to improve market design for electricity production in Australia, and improve our understanding of how information is incorporated into these markets.

Chapter 5

Conclusion

This thesis has examined the role of weather data in economic forecasts. A particular focus is the role of multi-horizon meteorological forecasts. Multi-horizon weather forecasts serve as a useful illustration of our broader research question: what can we learn from the revision structure of multi-horizon forecasts? Our focus on evaluating meteorological weather forecasts, and forecasts that use weather as an input, serves to highlight some of the challenges of exploiting The New Climate-Economy Literature for economic decision making.

In Chapter 2 we provided an illustration of how weather variables can be used to improve sales forecasts. Our findings illustrate how the uncertainty in meteorological weather forecasts may lead to a reduction in the accuracy of forecasts that use weather variables as predictors. The implication of this finding is that real-time data constraints are a concern for forecast models that use weather variables as an input. Specifically, the uncertainty contained in weather forecasts may dilute much of their explanatory power when used in economic forecasting models.

Chapter 3 outlines our proposed model of multi-horizon forecasts. We explain

how this model can be used as a unified framework for forecast evaluation. A particular focus of this chapter is explaining how existing test-based approaches to forecast evaluation nest as a special case of our approach, and as such our approach offers complementary insights. The approach to forecast evaluation we propose is flexible enough to detect multiple sources of forecasting error. These sources of error indicate the forecast information content and the forecast inefficiency at each horizon. Knowledge of horizon specific information content and forecast efficiency will be valuable for end users of forecasts, particularly when forecasts from multiple horizons may be utilised for decision making.

In Chapter 4 we apply the forecast evaluation approach developed in Chapter 3 to multi-horizon forecasts of electricity demand. The findings in this application illustrate that the forecast information content may depend not only on the horizon at which the forecasts are made, but also the time at which the forecasts are made. The relevance of the time of day is likely due to the infrequent updating of inputs to the forecasting model. The updating of inputs to a forecasting model, such as when newly available weather forecasts are used to update electricity demand forecasts, may occur at a lower frequency than the updating of the forecasts of interest. This insight is important for forecast users as well as those responsible for producing the forecasts. Forecasters who have a better understanding of the information content of their forecast revisions will be in a better position to evaluate the contribution of inputs to their forecasting model, particularly inputs that are infrequently updated.

Forecast evaluation is a multifaceted issue. This thesis makes two main contributions to the literature on forecast evaluation. First, we propose a methodology for illuminating multiple sources of forecasting error, and modelling the

structure of forecasting error along the forecast horizon. Second, we propose a novel measure of horizon specific forecast information content. Knowledge of how forecast information content evolves along the forecast horizon is important to forecast users, as this facilitates decision making at the optimal forecast horizon. Knowledge of horizon specific forecast information content is important for forecast practitioners, particularly when forecasts contain exogenous variables, as this knowledge may lead to a greater understanding of the predictability of the target variable at each forecast horizon.

The empirical focus of this thesis has been on the role of weather in economic forecasting. At the outset, we suggest that The New Climate-Economy Literature is ripe for exploitation by economic forecasters. Our results highlight some of the opportunities, and some of the challenges, of using weather forecasts for economic decision making. In conclusion, we suggest that the forecast evaluation methodology proposed in Chapter 3 of this thesis will have wider applications. We offer our model based approach to forecast evaluation as a method with which to examine the efficiency of all forecasts subject to later revision.

Appendix A

Selected Proofs

This appendix provides a formal connection between our model based approach to multi-horizon forecast evaluation and the variance covariance bound test approach to multi-horizon forecast evaluation proposed by Patton and Timmermann (2012). A description of the variables contained in our model is provided in Table A.1.

Table A.1: Definition of the multi-horizon forecast variables of interest

Variable	Definition
Target	$y_t \equiv \tilde{y}_t = \tilde{y}_{t-1} + \sigma_\varepsilon \eta_{\varepsilon,t} + \sum_{i=0}^{H-1} \sigma_{\nu_i} \eta_{i,t}$
Rational forecast	$\hat{y}^*_{t\|t-h} \equiv \tilde{y}_t + \nu_{t\|t-h} = \tilde{y}_{t-1} + \sigma_\varepsilon \eta_{\varepsilon,t} + \sum_{i=0}^{H-1} \sigma_{\nu_i} \eta_{i,t} - \sum_{i=0}^{h-1} \sigma_{\nu_i} \eta_{i,t}$
Rational-implicit forecast	$\hat{y}_{t\|t-h} \equiv \tilde{y}_t + \nu_{t\|t-h} + \zeta_{t\|t-h} = \tilde{y}_{t-1} + \sigma_\varepsilon \eta_{\varepsilon,t} + \sum_{i=0}^{H-1} \sigma_{\nu_i} \eta_{i,t} - \sum_{i=0}^{h-1} \sigma_{\nu_i} \eta_{i,t} + \sigma_{\zeta_h} \eta_{h,t}$
Rational forecasting error	$e^*_{t\|t-h} = y_t - \hat{y}^*_{t\|t-h} \equiv \sum_{i=0}^{h-1} \sigma_{\nu_i} \eta_{i,t}$
Rational-Implicit forecasting error	$e_{t\|t-h} = y_t - \hat{y}_{t\|t-h} \equiv \sum_{i=0}^{h-1} \sigma_{\nu_i} \eta_{i,t} - \sigma_{\zeta_h} \eta_{h,t}$
Rational revision between $h = m$ and $h = s$	$d^*_{t\|s,m} = \hat{y}^*_{t\|t-s} - \hat{y}^*_{t\|t-m} \equiv - \sum_{i=s}^{m-1} \sigma_{\nu_i} \eta_{i,t}$
Rational-Implicit revision between $h = m$ and $h = s$	$d_{t\|s,m} = \hat{y}_{t\|t-s} - \hat{y}_{t\|t-m} \equiv - \sum_{i=s}^{m-1} \sigma_{\nu_i} \eta_{i,t} - \sigma_{\zeta_m} \eta_{m,t} + \sigma_{\zeta_s} \eta_{s,t}$
Rational revision between $h = l$ and $h = s$	$d^*_{t\|s,l} = \hat{y}^*_{t\|t-s} - \hat{y}^*_{t\|t-l} \equiv - \sum_{i=s}^{l-1} \sigma_{\nu_i} \eta_{i,t}$
Rational-Implicit revision between $h = l$ and $h = s$	$d_{t\|s,l} = \hat{y}_{t\|t-s} - \hat{y}_{t\|t-l} \equiv - \sum_{i=s}^{l-1} \sigma_{\nu_i} \eta_{i,t} - \sigma_{\zeta_l} \eta_{l,t} + \sigma_{\zeta_s} \eta_{s,t}$

Notes: The generic $t - h$ forecast horizons considered above are $l > m > s$. H is the longest horizon at which forecasts are available.

A.1 Rational model

The following propositions show that multi-horizon forecasts that are characterised by the rational model satisfy all ten of the variance and covariance bound properties proposed by patton2012forecast.

Proposition 1 *The variance of rational forecasts should be non-increasing as the forecast horizon increases.*

Proof: Using the definition of a rational forecast from Table A.1 the variance of a rational forecast at horizon s is $\sigma_\varepsilon^2 + \sum_{i=s}^{H-1} \sigma_{\nu_i}^2$. Suppose we increase the forecast horizon from s to l. The variance of the rational forecast at horizon l is $\sigma_\varepsilon^2 + \sum_{i=l}^{H-1} \sigma_{\nu_i}^2 \leq \sigma_\varepsilon^2 + \sum_{i=s}^{H-1} \sigma_{\nu_i}^2$, for all $H \geq l > s$.

Proposition 2 *The variance of rational forecasting error should be non-decreasing as the forecast horizon increases.*

Proof: Using the definition of a rational forecasting error from Table A.1 the variance of rational forecasting error at horizon s is $\sum_{i=0}^{s-1} \sigma_{\nu_i}^2$. Suppose we increase the forecast horizon from s to l. The variance of rational forecasting error at horizon l is $\sum_{i=0}^{l-1} \sigma_{\nu_i}^2 \geq \sum_{i=0}^{s-1} \sigma_{\nu_i}^2$, for all $l > s$.

Proposition 3 *The variance of rational forecast revisions between a short horizon forecast and a long horizon forecast should be non-decreasing as the longer forecast horizon increases.*

Proof: Using the definition of a rational forecast revision from Table A.1 the variance of a rational forecast revision from horizon m to horizon s is $\sum_{i=s}^{m-1} \sigma_{\nu_i}^2$.

Suppose we increase the longer forecast horizon from m to l. The variance of a rational forecast revision from horizon l to horizon s is $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2 \geq \sum_{i=s}^{m-1} \sigma_{\nu_i}^2$, for all $l > m > s$.

Proposition 4 *The covariance between rational forecasts and the target variable should be non-increasing as the forecast horizon increases.*

Proof: Using the definition of a rational forecast and the definition of the target variable from Table A.1 the covariance between a rational forecast at horizon s and the target variable is $\sigma_{\varepsilon}^2 + \sum_{i=s}^{H-1} \sigma_{\nu_i}^2$. Suppose we increase the forecast horizon from s to l. The covariance between a rational forecast at horizon l and the target variable is $\sigma_{\varepsilon}^2 + \sum_{i=l}^{H-1} \sigma_{\nu_i}^2 \leq \sigma_{\varepsilon}^2 + \sum_{i=s}^{H-1} \sigma_{\nu_i}^2$, for all $H \geq l > s$.

Proposition 5 *The covariance between rational forecasts produced at two different forecast horizons should be non-increasing as the longer horizon increases.*

Proof: Using the definition of a rational forecast from Table A.1 the covariance between a rational forecast at horizon s and a rational forecast at horizon m is $\sigma_{\varepsilon}^2 + \sum_{i=m}^{H-1} \sigma_{\nu_i}^2$. Suppose we increase the longer forecast horizon from m to l. The covariance between a rational forecast at horizon s and a rational forecast at horizon l is $\sigma_{\varepsilon}^2 + \sum_{i=l}^{H-1} \sigma_{\nu_i}^2 \leq \sigma_{\varepsilon}^2 + \sum_{i=m}^{H-1} \sigma_{\nu_i}^2$, for all $H \geq l > m > s$.

Proposition 6 *The variance of a rational forecast revision between a short horizon forecast and a long horizon forecast should be bounded by twice the covariance between the target variable and the rational forecast revision between the short horizon forecast and long horizon forecast.*

Proof: Using the definition of a rational forecast revision and the definition of the target variable from Table A.1 the variance of a rational forecast revision from horizon l to horizon s is $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2$ and the covariance of the target variable and the rational forecast revision from horizon l to horizon s is $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2$. Therefore the condition $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2 \leq 2 \sum_{i=s}^{l-1} \sigma_{\nu_i}^2$ is trivially satisfied for all $l > s$.

Proposition 7 *The variance of a rational forecast revisions between a long horizon forecast and a medium horizon forecast should be bounded by twice the covariance between a rational short horizon forecast and the rational forecast revision between the long horizon forecast and medium horizon forecast.*

Proof: Using the definition of a rational forecast revision and the definition of a rational forecast from Table A.1 the variance of a rational forecast revision

between horizon l to horizon m is $\sum_{i=m}^{l-1} \sigma_{\nu_i}^2$ and the covariance between a rational forecast at horizon s and the rational forecast revision between horizon m and horizon s is $\sum_{i=m}^{l-1} \sigma_{\nu_i}^2$. Therefore the condition $\sum_{i=m}^{l-1} \sigma_{\nu_i}^2 \leq 2 \sum_{i=m}^{l-1} \sigma_{\nu_i}^2$ is trivially satisfied for all $l > m > s$.

Proposition 8 *The covariance between rational forecasting error and the target variable should be non-decreasing as the forecast horizon increases.*

Proof: Using the definition of rational forecasting error and the definition of the target variable from Table A.1 the covariance between the target variable and rational forecasting error at horizon s is $\sum_{i=0}^{s-1} \sigma_{\nu_i}^2$. Suppose we increase the forecast horizon from s to l. The covariance between the target variable and a rational forecasting error at horizon l is $\sum_{i=0}^{l-1} \sigma_{\nu_i}^2 \geq \sum_{i=0}^{s-1} \sigma_{\nu_i}^2$, for all $l > s$.

Proposition 9 *The covariance between a rational short horizon forecast and the rational forecast revision between a longer horizon forecast and the short horizon forecast should be non-decreasing as the longer forecast horizon increases.*

Proof: Using the definition of a rational forecast revision and the definition of a rational forecast from Table A.1 the covariance between a rational forecast at horizon s and a rational forecast revision between horizon m and horizon s is $\sum_{i=s}^{m-1} \sigma_{\nu_i}^2$. Suppose we increase the longer forecast horizon from m to l. The covariance between a rational forecast at horizon s and a rational forecast revision between horizon l and horizon s is $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2 \geq \sum_{i=s}^{m-1} \sigma_{\nu_i}^2$, for all $l > m > s$.

Proposition 10 *The covariance between forecasting error from a long horizon forecast and the rational forecast revision between a long horizon forecast and a short horizon forecast should be non-decreasing as the longer forecast horizon increases.*

Proof: Using the definition of a rational forecast revision and the definition of a rational forecasting error from Table A.1 the covariance between a rational forecasting error from horizon m and a rational forecast revision between horizon m and horizon s is $\sum_{i=s}^{m-1} \sigma_{\nu_i}^2$. Suppose we increase the longer forecast horizon from m to l. The covariance between a rational forecasting error at horizon l and the revision between horizon l and horizon s is $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2 \geq \sum_{i=s}^{m-1} \sigma_{\nu_i}^2$, for all $l > m > s$.

A.2 Rational-Implicit model

The following propositions show the conditions under which multi-horizon forecasts that are characterised by the rational-implicit model may satisfy the variance and covariance bound properties proposed by patton2012forecast.

Proposition 11 *Under certain conditions the variance of rational-implicit forecasts should be non-increasing as the forecast horizon increases.*

Proof: Using the definition of a rational-implicit forecast from Table A.1 the variance of a rational-implicit forecast at horizon s is $\sigma_\varepsilon + \sum_{i=s}^{H-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_s}^2$. Suppose we increase the forecast horizon from s to l. The variance of the rational-implicit forecast at horizon l is $\sigma_\varepsilon^2 + \sum_{i=l}^{H-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_l}^2 \leq \sigma_\varepsilon^2 + \sum_{i=s}^{H-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_s}^2$, if and only if $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2 \geq \sigma_{\zeta_l}^2 - \sigma_{\zeta_s}^2$, for all $l > s$.

Proposition 12 *Under certain conditions the variance of rational-implicit forecasting error should be non-decreasing as the forecast horizon increases.*

Proof: Using the definition of a rational-implicit forecasting error from Table A.1 the variance of rational-implicit forecasting error at horizon s is $\sum_{i=0}^{s-1}\sigma_{\nu_i}^2+\sigma_{\zeta_s}^2$. Suppose we increase the forecast horizon from s to l. The variance of the rational-implicit forecasting error at horizon l is $\sum_{i=0}^{l-1}\sigma_{\nu_i}^2+\sigma_{\zeta_l}^2 \geq \sum_{i=0}^{s-1}\sigma_{\nu_i}^2+\sigma_{\zeta_s}^2$, if and only if $\sum_{i=s}^{l-1}\sigma_{\nu_i}^2 \geq \sigma_{\zeta_s}^2-\sigma_{\zeta_l}^2$, for all $l>s$.

Proposition 13 *Under certain conditions the variance of a rational-implicit forecast revision between a short horizon forecast and a long horizon forecast should be non-decreasing as the longer forecast horizon increases.*

Proof: Using the definition of a rational-implicit forecast revision from Table A.1 the variance of a rational-implicit forecast revision between horizon m and horizon s is $\sum_{i=s}^{m-1}\sigma_{\nu_i}^2+\sigma_{\zeta_m}^2+\sigma_{\zeta_s}^2$. Suppose we increase the longer forecast horizon from m to l. The variance of a rational-implicit forecast revision between horizon l and horizon s is $\sum_{i=s}^{l-1}\sigma_{\nu_i}^2+\sigma_{\zeta_l}^2+\sigma_{\zeta_s}^2 \geq \sum_{i=s}^{m-1}\sigma_{\nu_i}^2+\sigma_{\zeta_m}^2+\sigma_{\zeta_s}^2$, if and only if $\sum_{i=m}^{l-1}\sigma_{\nu_i}^2 \geq \sigma_{\zeta_m}^2-\sigma_{\zeta_l}^2$, for all $l>m>s$.

Proposition 14 *The covariance between rational-implicit forecasts and the target variable should be non-increasing as the forecast horizon increases.*

Proof: Using the definition of a rational-implicit forecast and the definition of the target variable from Table A.1 the covariance between a rational-implicit forecast at horizon s and the target variable is $\sigma_{\varepsilon}^2+\sum_{i=s}^{H-1}\sigma_{\nu_i}^2$. Suppose we

increase the forecast horizon from s to l. The covariance between a rational-implicit forecast at horizon l and the target variable is $\sigma_\varepsilon^2 + \sum_{i=l}^{H-1} \sigma_{\nu_i}^2 \leq \sigma_\varepsilon^2 + \sum_{i=s}^{H-1} \sigma_{\nu_i}^2$, for all $l > s$.

Proposition 15 *The covariance between rational-implicit forecasts produced at a short horizon and a long horizon should be non-increasing as the longer horizon increases.*

Proof: Using the definition of a rational-implicit forecast from Table A.1 the covariance between a rational-implicit forecast at horizon s and a rational-implicit forecast at horizon m is $\sigma_\varepsilon^2 + \sum_{i=m}^{H-1} \sigma_{\nu_i}^2$. Suppose we increase the longer forecast horizon from m to l. The covariance between a rational-implicit forecast at horizon s and a rational-implicit forecast at horizon l is $\sigma_\varepsilon^2 + \sum_{i=l}^{H-1} \sigma_{\nu_i}^2 \leq \sigma_\varepsilon^2 + \sum_{i=m}^{H-1} \sigma_{\nu_i}^2$, for all $l > m > s$.

Proposition 16 *Under certain conditions the variance of a rational-implicit forecast revision between a long horizon forecast and a short horizon forecast may be bounded by twice the covariance between the target variable and the rational-implicit forecast revision between the long horizon forecast and short horizon forecast.*

Proof: Using the definition of a rational-implicit forecast revision and the definition of the target variable from Table A.1 the variance of a rational-implicit forecast revision between horizon l and horizon s is $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_s}^2 + \sigma_{\zeta_l}^2$ and the covariance betweeen the target variable and the rational-implicit forecast revision between horizon l and horizon s is $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2$. Therefore the condition $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_s}^2 + \sigma_{\zeta_l}^2 \leq 2 \sum_{i=s}^{l-1} \sigma_{\nu_i}^2$ is satisfied if and only if $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2 \geq \sigma_{\zeta_s}^2 + \sigma_{\zeta_l}^2$, for all $l > s$ where s is fixed.

Proposition 17 *Under certain conditions the variance of a rational-implicit forecast revision between a long horizon forecast and a medium horizon forecast should be bounded by twice the covariance between a rational-implicit short horizon forecast and the rational-implicit forecast revision between the long horizon forecast and medium horizon forecast.*

Proof: Using the definition of a rational-implicit forecast revision and the definition of a rational-implicit forecast from Table A.1 the variance of a rational-implicit forecast revision between horizon l and horizon m is $\sum_{i=m}^{l-1}\sigma_{\nu_i}^2+\sigma_{\zeta_m}^2+\sigma_{\zeta_l}^2$ and the covariance between a rational-implicit forecast at horizon s and the rational-implicit forecast revision between horizon m and horizon s is $\sum_{i=m}^{l-1}\sigma_{\nu_i}^2$. Therefore the condition $\sum_{i=m}^{l-1}\sigma_{\nu_i}^2+\sigma_{\zeta_m}^2+\sigma_{\zeta_l}^2\leq 2\sum_{i=m}^{l-1}\sigma_{\nu_i}^2$ is satisfied if and only if $\sum_{i=m}^{l-1}\sigma_{\nu_i}^2\geq\sigma_{\zeta_m}^2+\sigma_{\zeta_l}^2$, for all $l>m$ where m is fixed.

Proposition 18 *The covariance between rational-implicit forecasting error and the target variable should be non-decreasing as the forecast horizon increases.*

Proof: Using the definition of rational-implicit forecasting error and the definition of the target variable from Table A.1 the covariance between the target variable and rational-implicit forecasting error at horizon s is $\sum_{i=0}^{s-1}\sigma_{\nu_i}^2$. Suppose we increase the forecast horizon from s to l. The covariance between rational-implicit forecasting error and the target variable at horizon l is $\sum_{i=0}^{l-1}\sigma_{\nu_i}^2\geq\sum_{i=0}^{s-1}\sigma_{\nu_i}^2$, for all $l>s$.

Proposition 19 *The covariance between a rational-implicit short horizon forecast and the rational-implicit forecast revision between a longer horizon forecast*

and the short horizon forecast should be non-decreasing as the longer forecast horizon increases.

Proof: Using the definition of a rational-implicit forecast revision and the definition of a rational-implicit forecast from Table A.1 the covariance between a rational-implicit forecast at horizon s and a rational-implicit forecast revision between horizon m and horizon s is $\sum_{i=s}^{m-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_s}^2$. Suppose we increase the longer forecast horizon from m to l. The covariance between a rational-implicit forecast at horizon s and a rational-implicit forecast revision between horizon l and horizon s is $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_s}^2 \geq \sum_{i=s}^{m-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_s}^2$, for all $l > m > s$.

Proposition 20 *Under certain conditions the covariance between a long horizon rational-implicit forecasting error and the rational-implicit forecast revision between the long horizon forecast and a short horizon forecast should be non-decreasing as the longer forecast horizon increases.*

Proof: Using the definition of a rational-implicit forecast revision and the definition of a rational-implicit forecasting error from Table A.1 the covariance between a rational-implicit forecasting error from horizon m and a rational-implicit forecast revision between horizon m and horizon s is $\sum_{i=s}^{m-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_m}^2$. Suppose we increase the longer forecast horizon from m to l. The covariance between a rational-implicit forecasting error from horizon l and a rational-implicit revision between horizon l and horizon s is $\sum_{i=s}^{l-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_l}^2 \geq \sum_{i=s}^{m-1} \sigma_{\nu_i}^2 + \sigma_{\zeta_m}^2$, if and only if $\sum_{i=m}^{l-1} \sigma_{\nu_i}^2 \geq \sigma_{\zeta_m}^2 - \sigma_{\zeta_l}^2$, for all $l > m > s$.

Appendix B

Additional Figures & Tables

B.1

Table B.1: Parameter estimates for sub-samples of the maximum daily temperature forecast series, Melbourne Australia.

		Panel A				Panel B			
		Feb 1, 2009 to May 21, 2012		**May 22, 2012 to Dec 31, 2014**		**Summer**		**Winter**	
Parameters		Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.
Rational Revisions	$\sigma_{\omega_{13}}$	0.03	(0.02)	0.04	(0.03)	0.01	(0.44)	0.04	(0.01)
	$\sigma_{\omega_{12}}$	0.45	(0.05)	1.10	(0.09)	1.05	(0.08)	0.47	(0.05)
	$\sigma_{\omega_{11}}$	0.44	(0.04)	0.80	(0.11)	0.81	(0.10)	0.36	(0.05)
	$\sigma_{\omega_{10}}$	0.48	(0.04)	0.81	(0.10)	0.85	(0.09)	0.37	(0.05)
	σ_{ω_9}	0.49	(0.04)	0.85	(0.09)	0.86	(0.08)	0.42	(0.05)
	σ_{ω_8}	0.64	(0.04)	1.30	(0.07)	1.26	(0.07)	0.63	(0.04)
	σ_{ω_7}	0.93	(0.04)	1.67	(0.06)	1.63	(0.06)	0.90	(0.09)
	σ_{ω_6}	3.56	(0.11)	3.01	(0.09)	4.11	(0.11)	2.34	(0.07)
Implicit Errors	$\sigma_{\zeta_{14}}$	0.81	(0.03)	1.54	(0.06)	1.55	(0.06)	0.71	(0.03)
	$\sigma_{\zeta_{13}}$	0.75	(0.03)	1.37	(0.06)	1.39	(0.05)	0.66	(0.03)
	$\sigma_{\zeta_{12}}$	0.74	(0.02)	1.35	(0.06)	1.34	(0.06)	0.69	(0.03)
	$\sigma_{\zeta_{11}}$	0.72	(0.03)	1.33	(0.06)	1.31	(0.05)	0.68	(0.03)
	$\sigma_{\zeta_{10}}$	0.69	(0.03)	1.31	(0.06)	1.28	(0.05)	0.67	(0.03)
	σ_{ζ_9}	0.59	(0.03)	1.14	(0.06)	1.11	(0.06)	0.57	(0.03)
	σ_{ζ_8}	0.46	(0.04)	0.62	(0.11)	0.67	(0.09)	0.37	(0.05)
	σ_{ζ_7}	0.02	(0.01)	0.04	(0.06)	0.06	(0.04)	0.08	(0.89)
Akaike Info Criterion				74,244				73,372	
Logarithmic likelihood				-37,089				-36,653	

		Panel C							
		Feb 1, 2009 to May 21, 2012				**May 22, 2012 to Dec 31, 2014**			
		Summer		**Winter**		**Summer**		**Winter**	
Parameters		Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.
Rational Revisions	$\sigma_{\omega_{13}}$	0.08	(0.24)	0.04	(0.02)	0.02	(0.15)	0.05	(0.02)
	$\sigma_{\omega_{12}}$	0.54	(0.07)	0.32	(0.04)	1.41	(0.13)	0.60	(0.08)
	$\sigma_{\omega_{11}}$	0.56	(0.06)	0.29	(0.05)	1.03	(0.16)	0.43	(0.09)
	$\sigma_{\omega_{10}}$	0.59	(0.07)	0.34	(0.04)	1.06	(0.15)	0.40	(0.09)
	σ_{ω_9}	0.61	(0.06)	0.33	(0.04)	1.10	(0.14)	0.51	(0.08)
	σ_{ω_8}	0.80	(0.05)	0.42	(0.04)	1.66	(0.11)	0.83	(0.06)
	σ_{ω_7}	1.10	(0.06)	0.72	(0.04)	2.11	(0.09)	1.09	(0.06)
	σ_{ω_6}	4.42	(0.16)	2.45	(0.09)	3.69	(0.14)	2.18	(0.08)
Implicit Errors	$\sigma_{\zeta_{14}}$	1.01	(0.04)	0.55	(0.03)	2.04	(0.09)	0.88	(0.05)
	$\sigma_{\zeta_{13}}$	0.97	(0.04)	0.47	(0.02)	1.83	(0.09)	0.83	(0.05)
	$\sigma_{\zeta_{12}}$	0.92	(0.04)	0.53	(0.02)	1.76	(0.09)	0.85	(0.04)
	$\sigma_{\zeta_{11}}$	0.89	(0.04)	0.50	(0.03)	1.71	(0.09)	0.85	(0.05)
	$\sigma_{\zeta_{10}}$	0.86	(0.04)	0.49	(0.03)	1.67	(0.09)	0.84	(0.05)
	σ_{ζ_9}	0.74	(0.04)	0.42	(0.03)	1.46	(0.09)	0.72	(0.05)
	σ_{ζ_8}	0.56	(0.06)	0.33	(0.04)	0.79	(0.17)	0.42	(0.09)
	σ_{ζ_7}	0.07	(0.03)	0.06	(0.06)	0.13	(0.12)	0.12	(0.26)
Akaike Info Criterion				70,931					
Logarithmic likelihood				-35,400					

Notes to Figure B.1: Standard errors are in parentheses. Data are from Stern and Davidson (2015) and include $t = 2,160$ maximum daily temperature observations (degrees Celsius) and $q \times t = 14 \times 2,160$ meteorological forecasts of maximum daily temperature produced at horizons $h = 1$ day to $h = 14$ days out from the observation date. The forecasts $\hat{y}_{t|t-h}$ are bias adjusted, where the bias follows a local linear trend process. The three alternative models of multi-horizon forecasts we consider are rational-implicit models ($\hat{y}_{t|t-h} = \tilde{y}_t + \nu_{t|t-h} + \zeta_{t|t-h}$) with time varying parameters. The time varying parameter models we consider are: Panel A) separate parameters before and after the upgrade of a new Numerical Weather Prediction model on May 22, 2012, used as an input to the forecast generation process; Panel B) separate parameters for summer months, assumed to begin on September 21, and winter months assumed to begin on March 21, roughly corresponding with spring and autumn equinoxes, respectively; Panel C) separate parameters for summer months and winter months before and after May 22, 2012. Comparing AIC values for these three models we conclude that the data is best characterised by the rational-implicit model with separate parameters for summer months and winter months before and after May 22, 2012. Parameter σ_{ω_h} is the standard deviation of the rational forecast revision between horizon $h+1$ and horizon h. Parameter σ_{ζ_h} is the standard deviation of the implicit forecasting error component at horizon h. During the summer months parameter estimates σ_{ω_h} show higher variability in rational forecast revisions relative to variability in rational forecast revisions during the winter months. We interpret this parameter as the increase in information content due to forecast revisions. Higher variability in the rational forecast revisions over the summer months reflects higher variability in the maximum daily temperature observation series over the summer months. Parameter estimates σ_{ω_7} to $\sigma_{\omega_{12}}$ during both summer and winter months show a higher rate of information accumulation after an upgrade to one of the models used as an input to the forecasts on May 22, 2012, relative to the period before this upgrade occurred. Parameter estimates σ_{ζ_7} to $\sigma_{\zeta_{12}}$ also increase after May 22, 2012 indicating an increase the amount of implicit error in the forecasts at these horizons.

Appendix C

MAPE(%) HEAT MAPS OF DEMAND FORECAST PERFORMANCE, BY FORECAST HORIZON

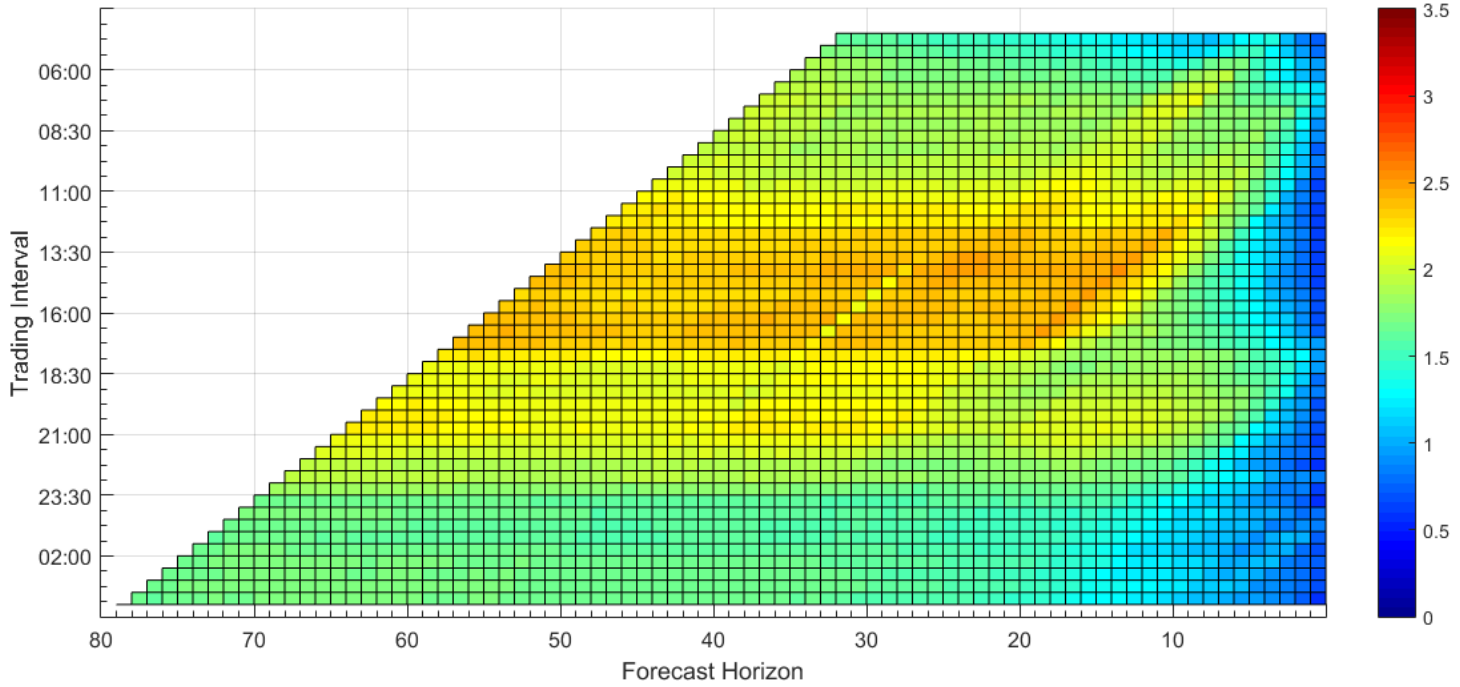


Figure C.1: Monday electricity trading day: — MAPE(%) heat map of demand forecast performance, by forecast horizon.

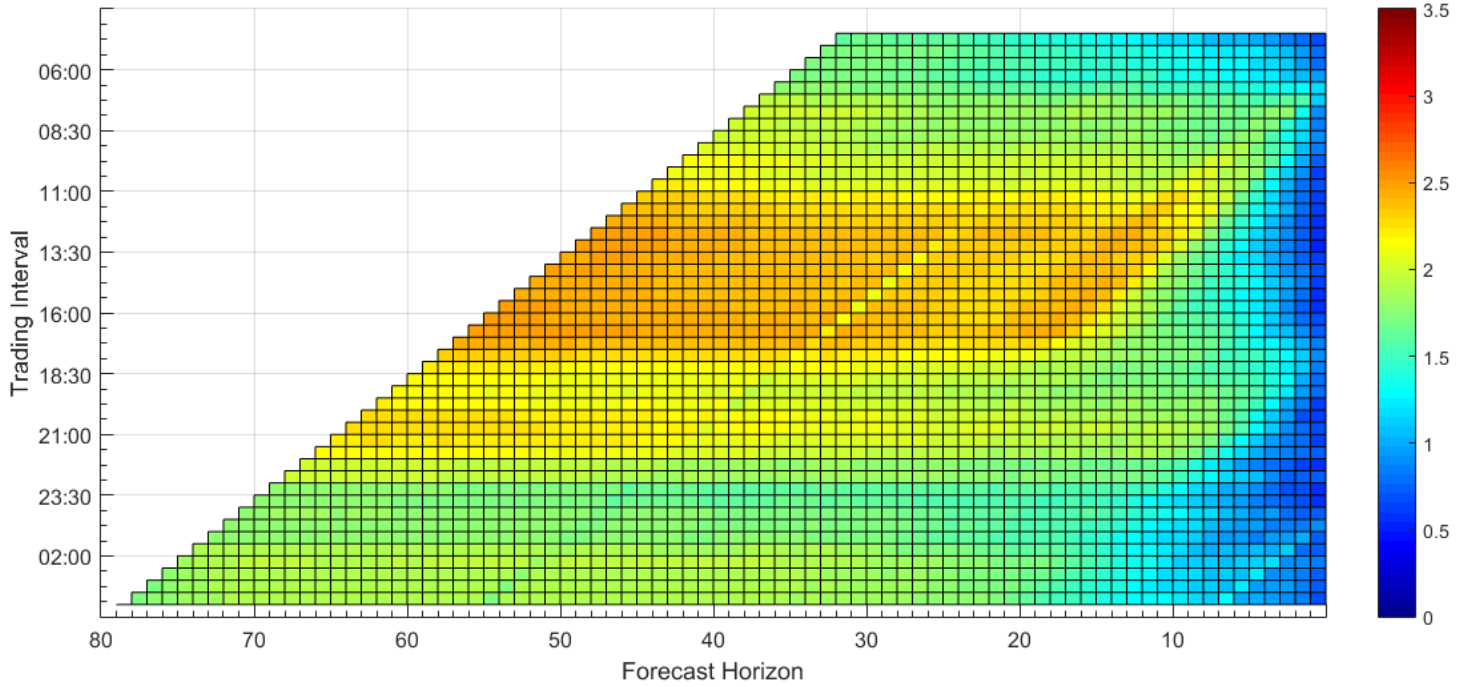


Figure C.2: Tuesday electricity trading day — MAPE(%) heat map of demand forecast performance, by forecast horizon.

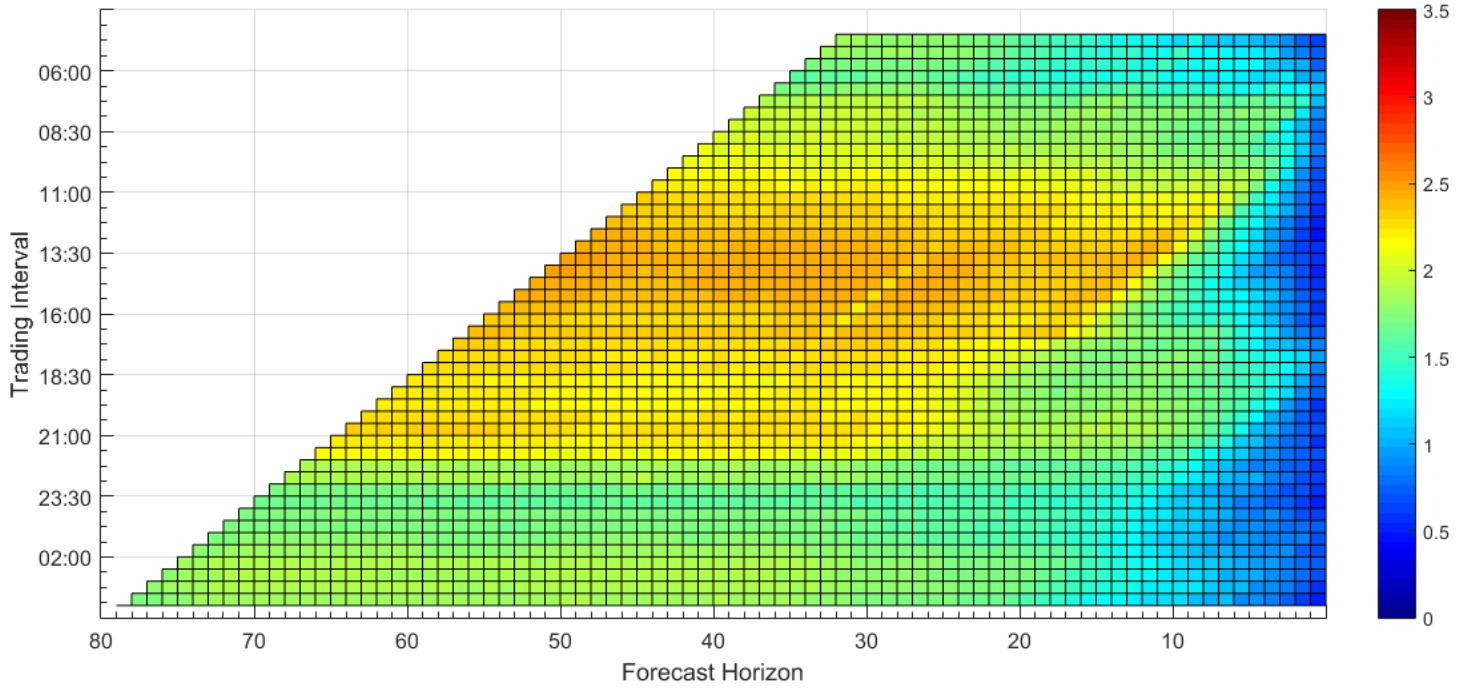


Figure C.3: Wednesday electricity trading day — MAPE(%) heat map of demand forecast performance, by forecast horizon.

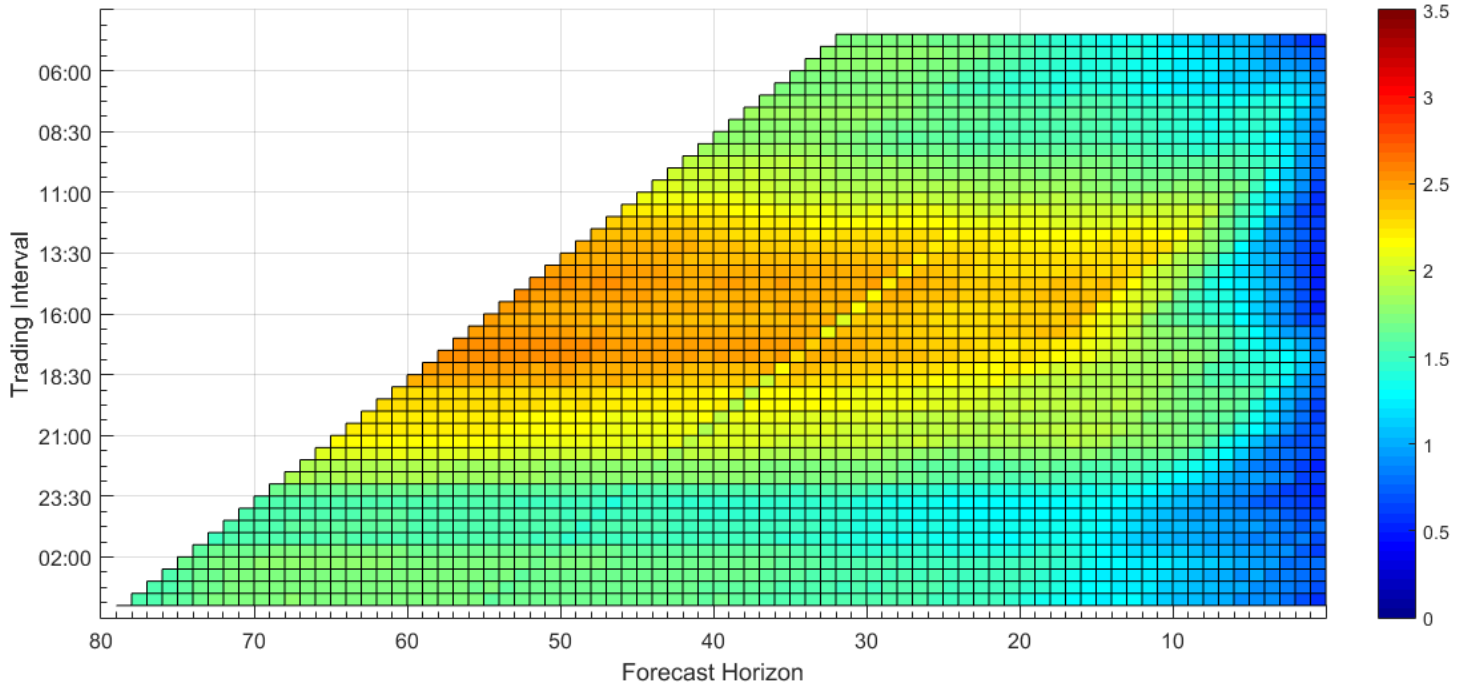


Figure C.4: Thursday electricity trading day — MAPE(%) heat map of demand forecast performance, by forecast horizon.

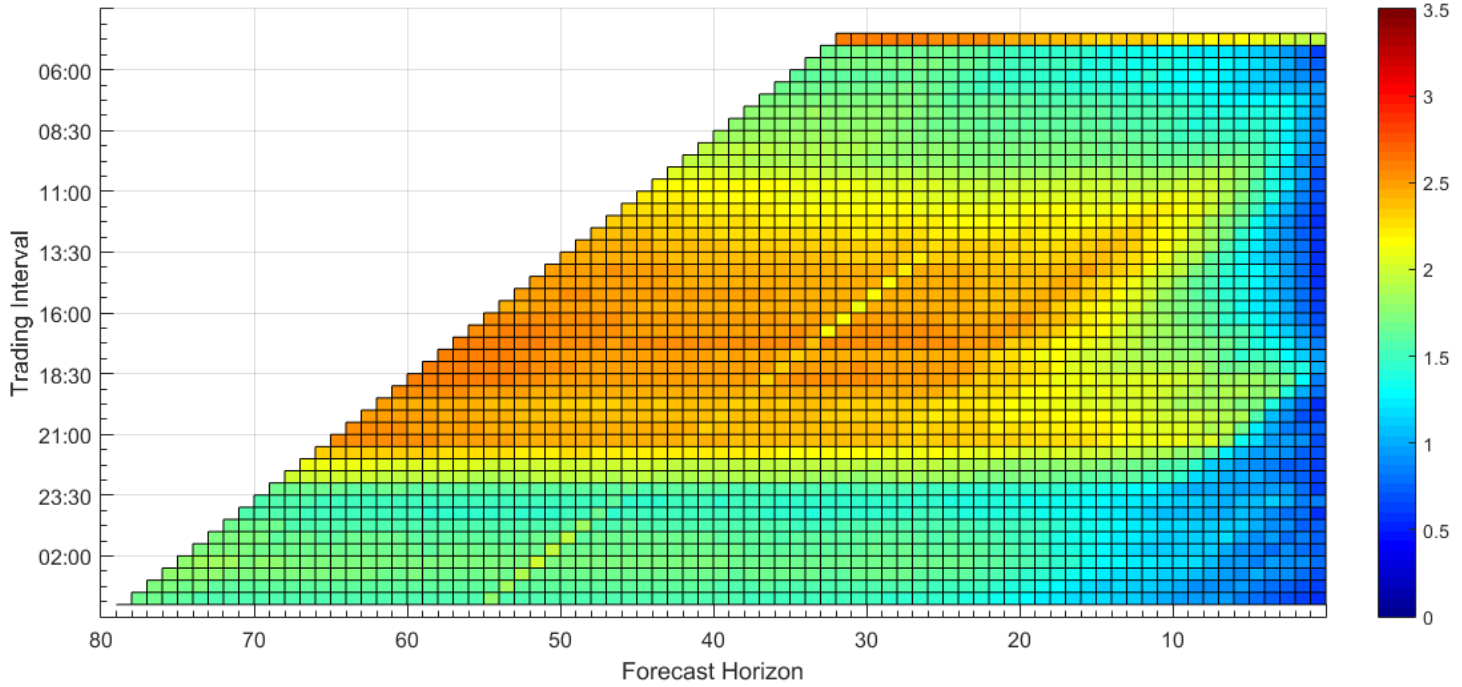


Figure C.5: Friday electricity trading day — MAPE(%) heat map of demand forecast performance, by forecast horizon.

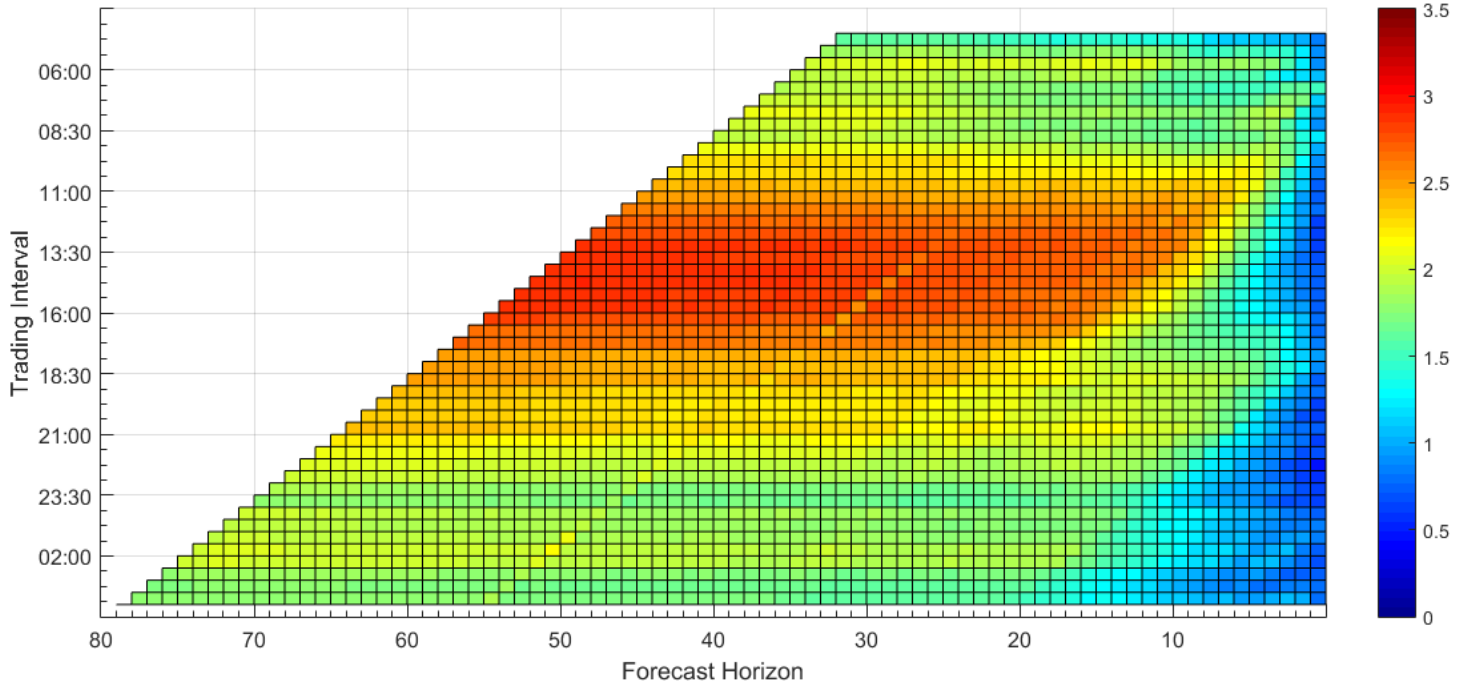


Figure C.6: Saturday electricity trading day — MAPE(%) heat map of demand forecast performance, by forecast horizon.

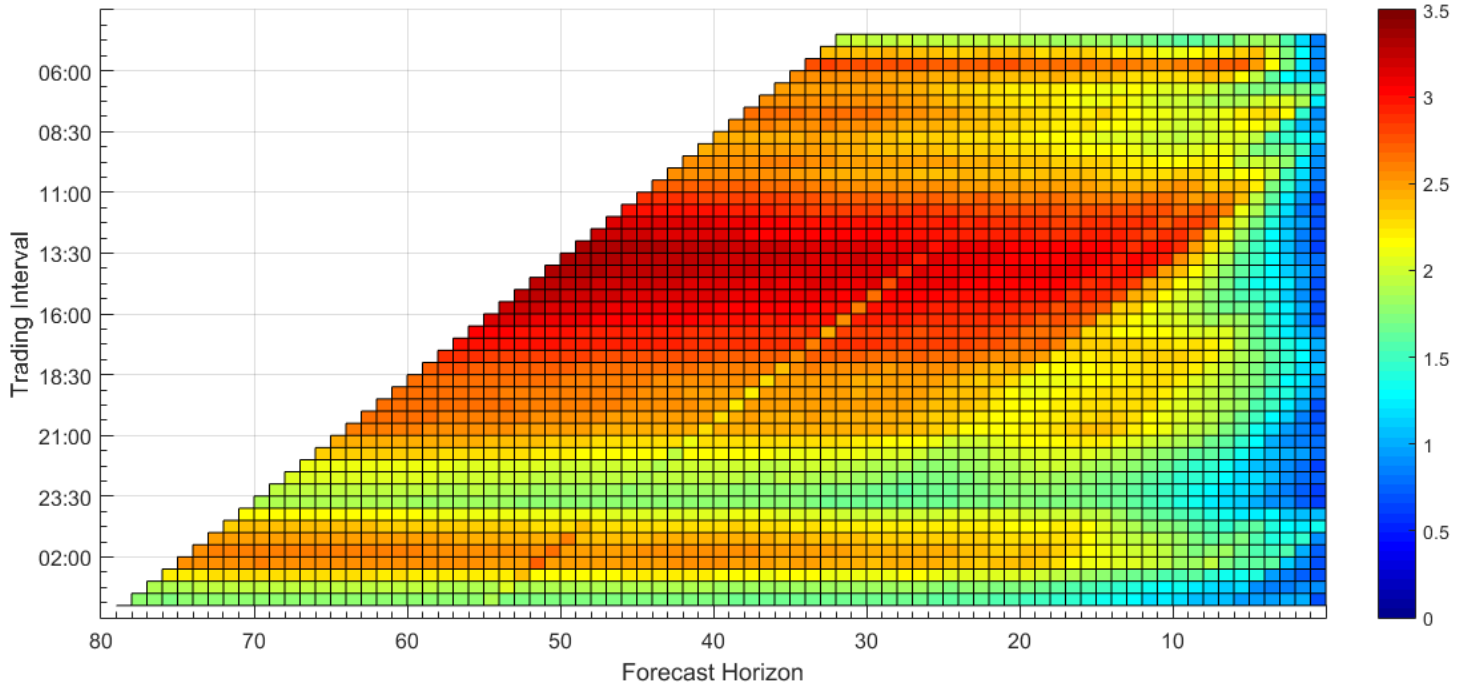


Figure C.7: Sunday electricity trading day — MAPE(%) heat map of demand forecast performance, by forecast horizon.

Appendix D

ESTIMATES OF BIAS IN ELECTRICITY DEMAND FORECASTS

Figure D.1: Monday electricity trading day, 04:30 to 10:00 — estimates of bias in the demand forecasts.

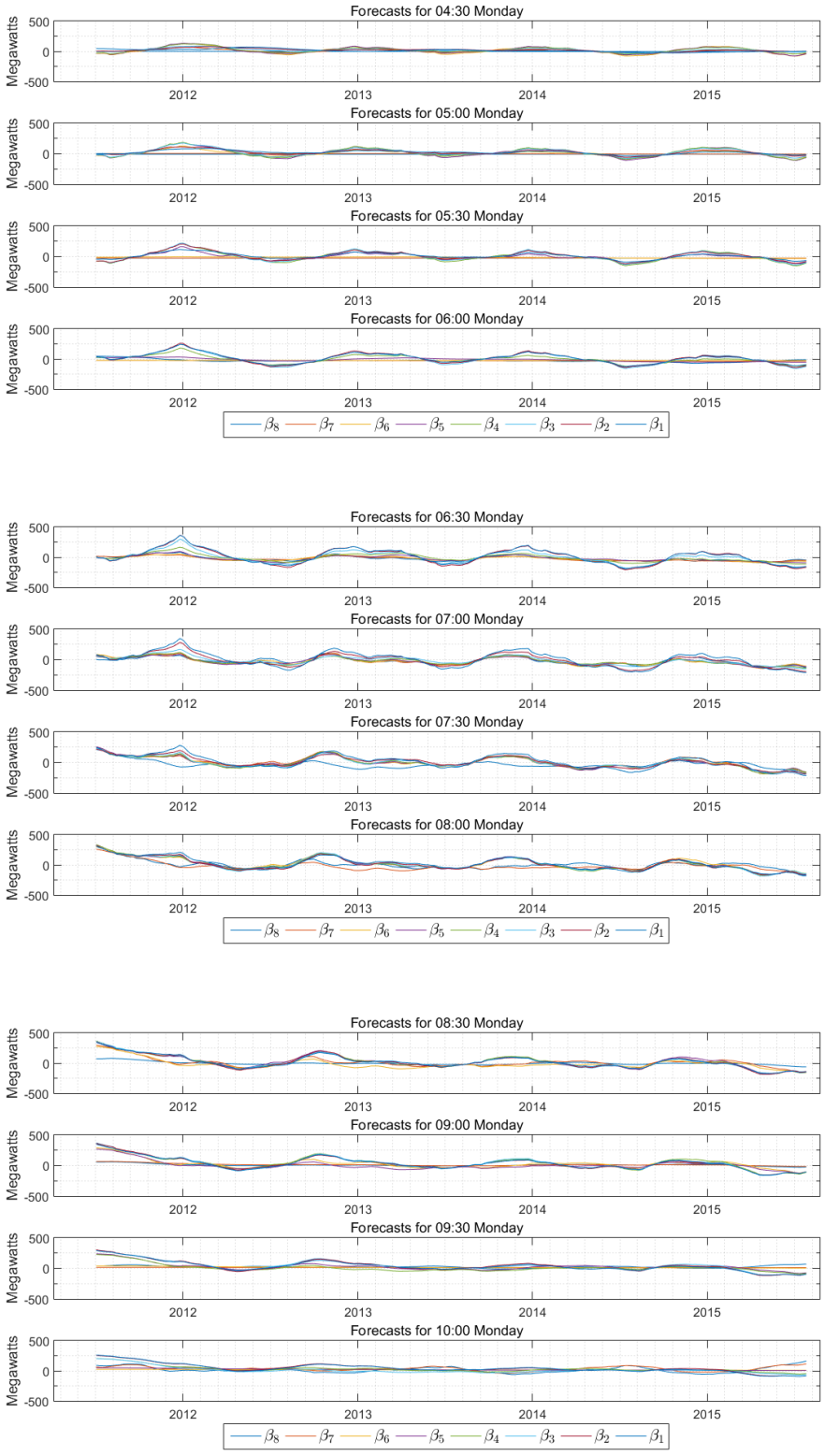


Figure D.2: Monday electricity trading day, 10:30 to 16:00 — estimates of bias in the demand forecasts.

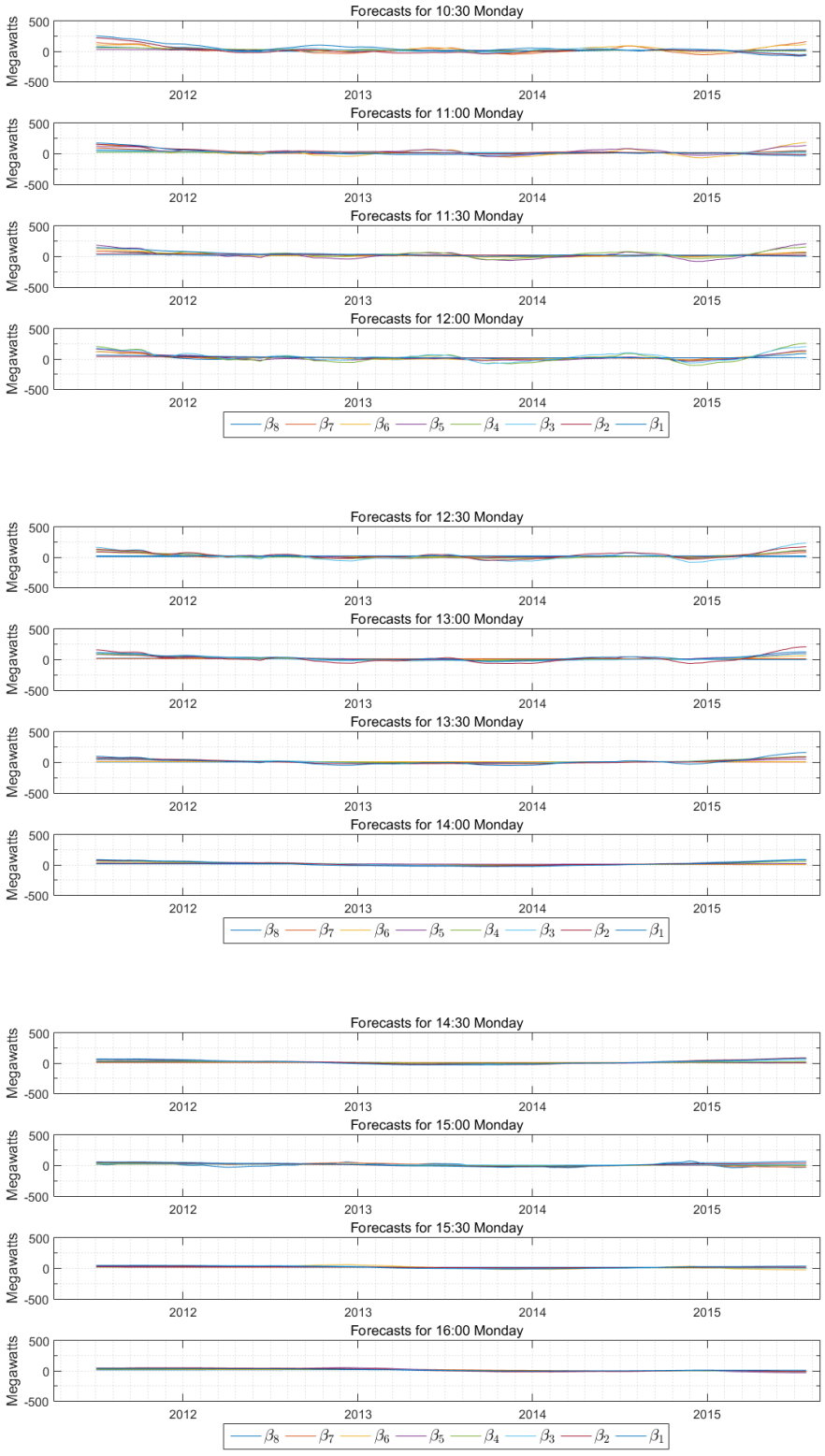


Figure D.3: Monday electricity trading day, 16:30 to 22:00 — estimates of bias in the demand forecasts.

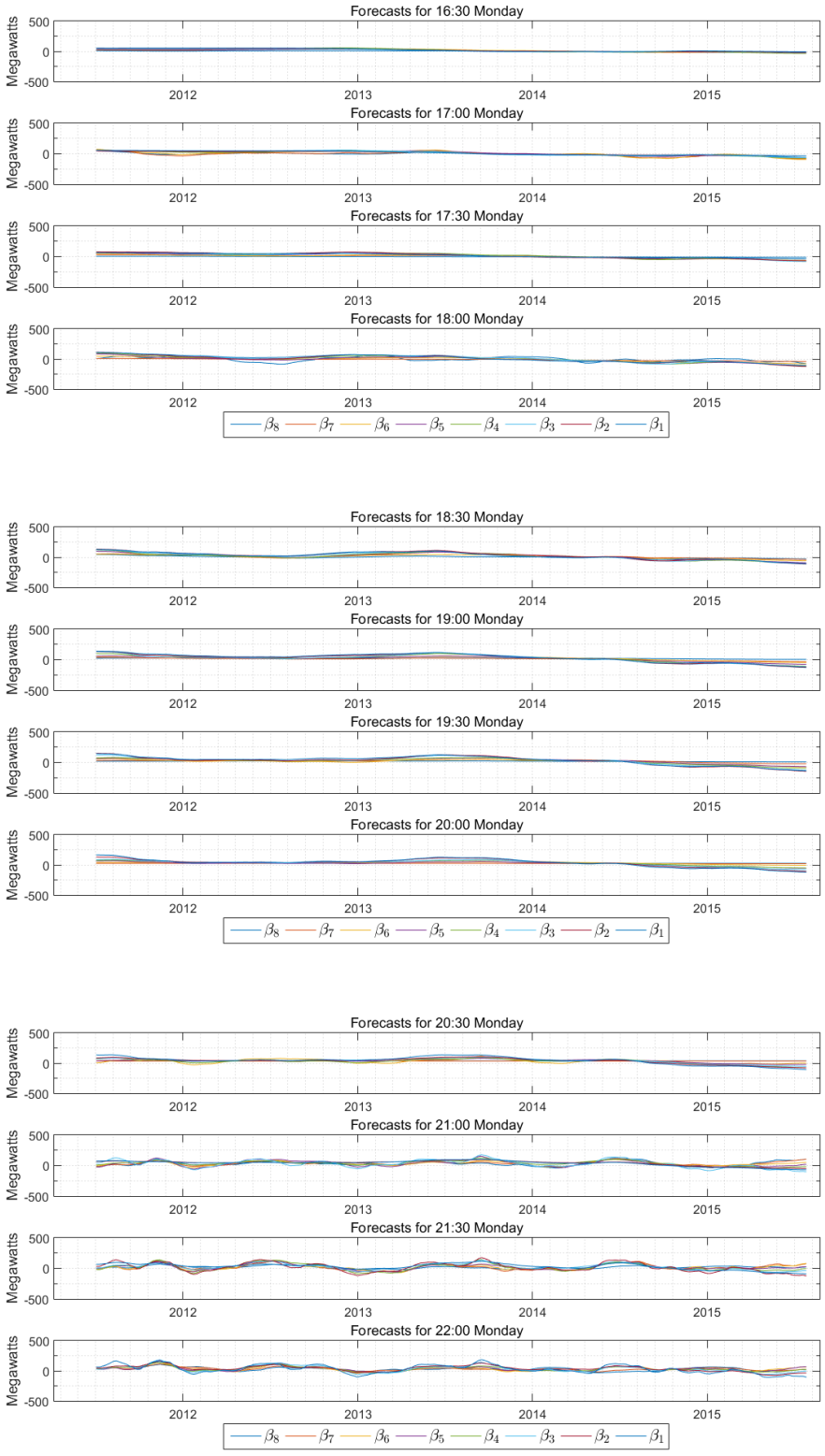


Figure D.4: Monday electricity trading day, 22:30 to 04:00 — estimates of bias in the demand forecasts.

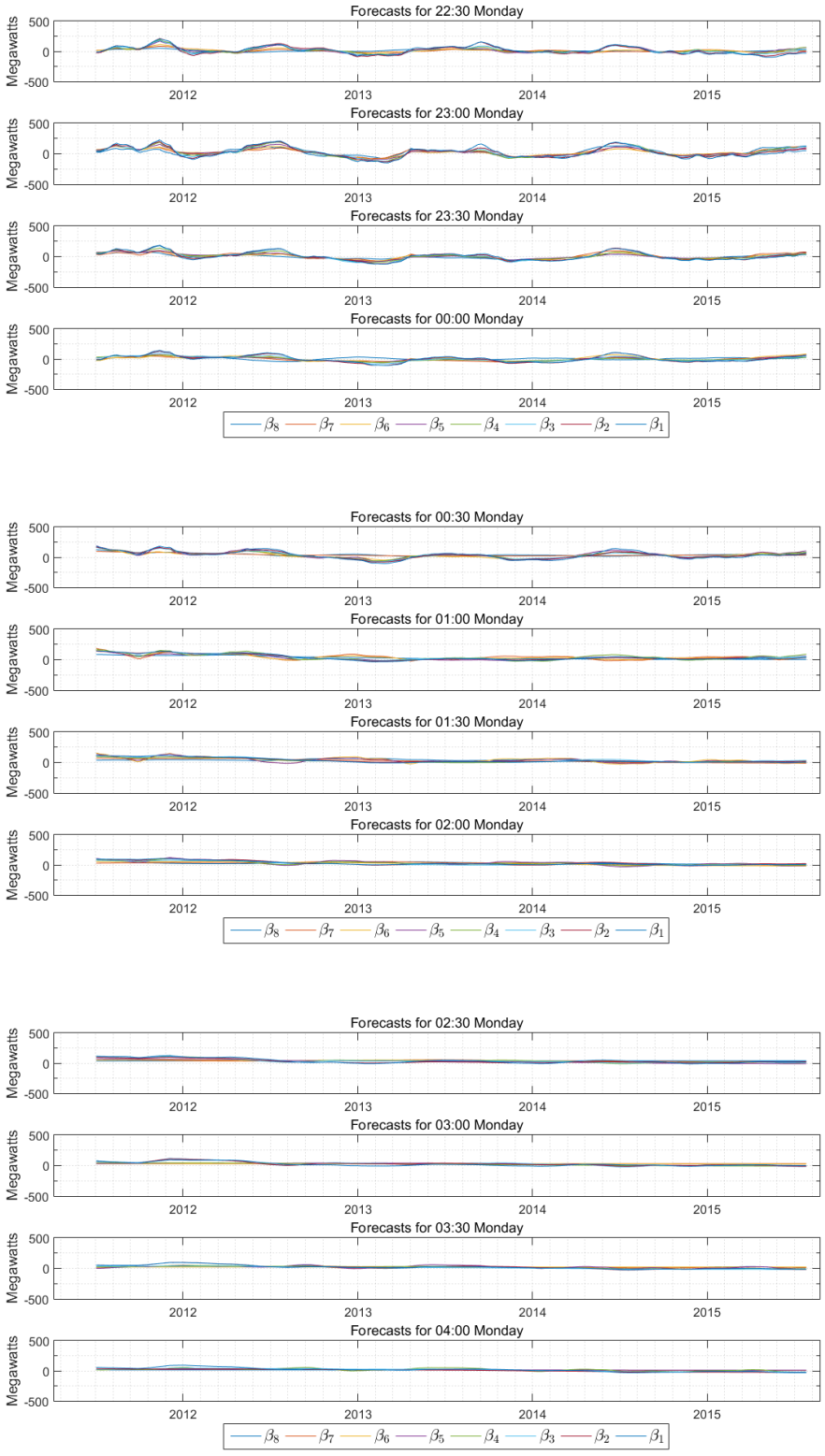


Figure D.5: Tuesday electricity trading day, 04:30 to 10:00 — estimates of bias in the demand forecasts.

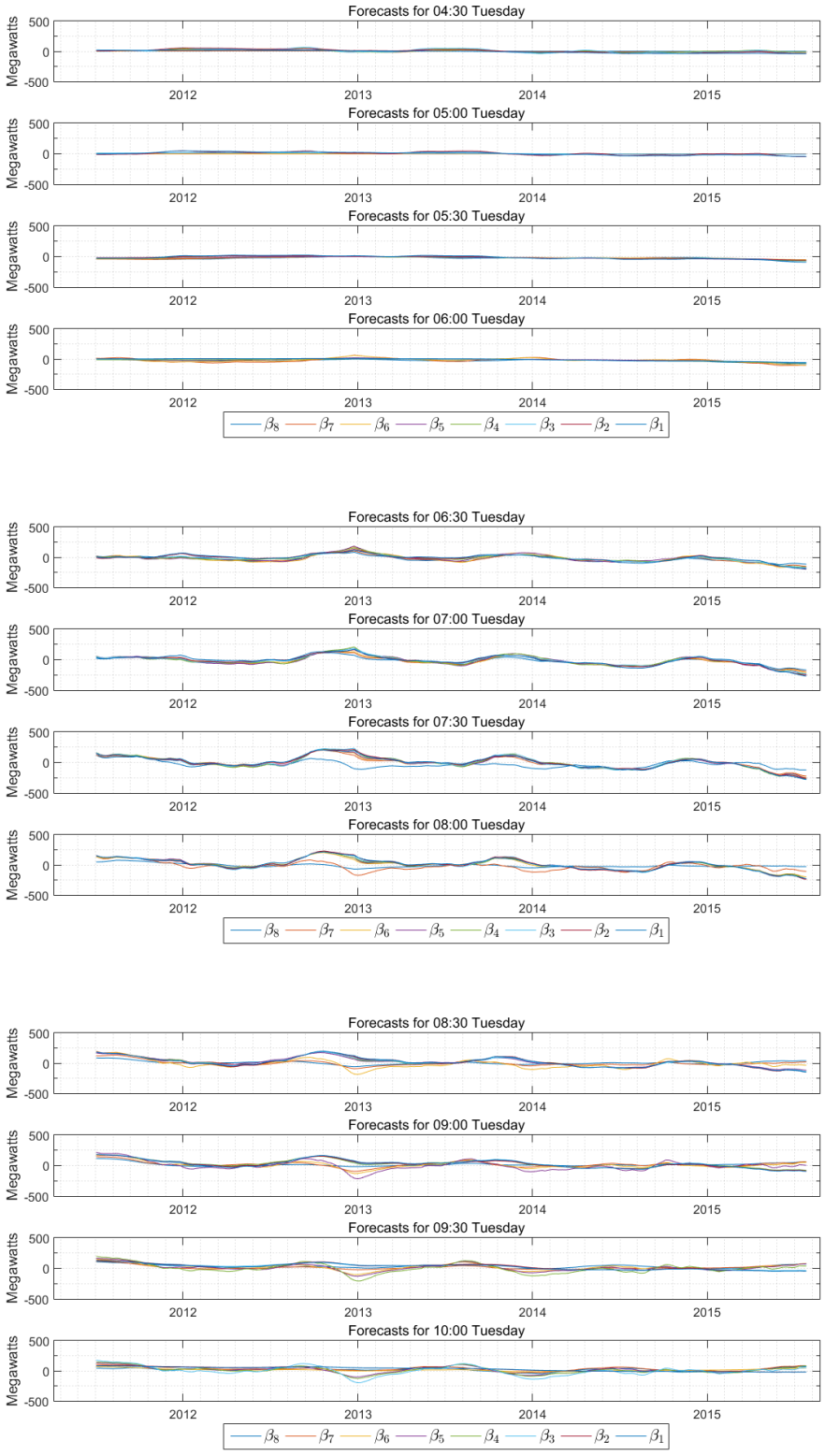


Figure D.6: Tuesday electricity trading day, 10:30 to 16:00 — estimates of bias in the demand forecasts.

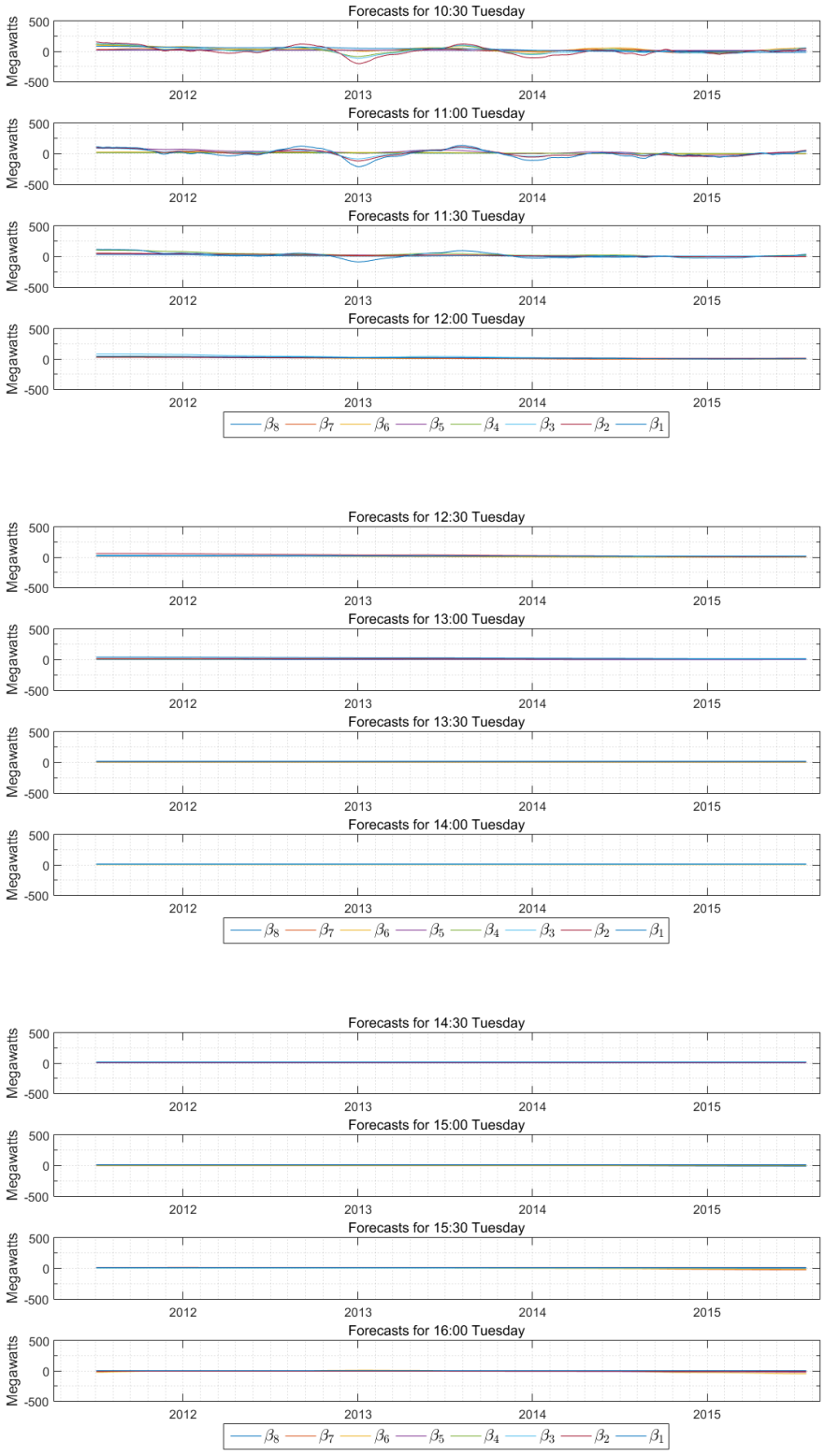


Figure D.7: Tuesday electricity trading day, 16:30 to 22:00 — estimates of bias in the demand forecasts.

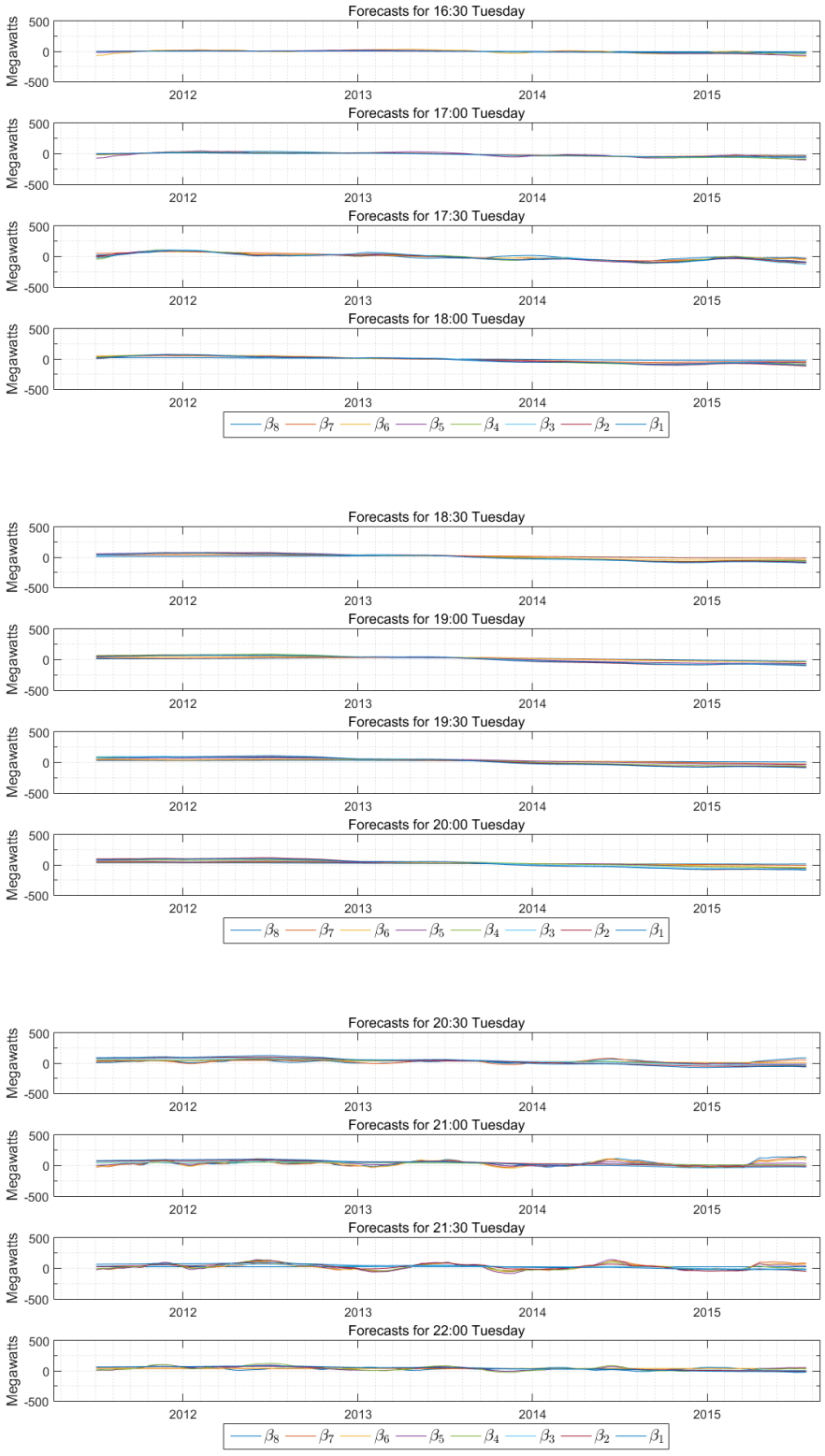


Figure D.8: Tuesday electricity trading day, 22:30 to 04:00 — estimates of bias in the demand forecasts.

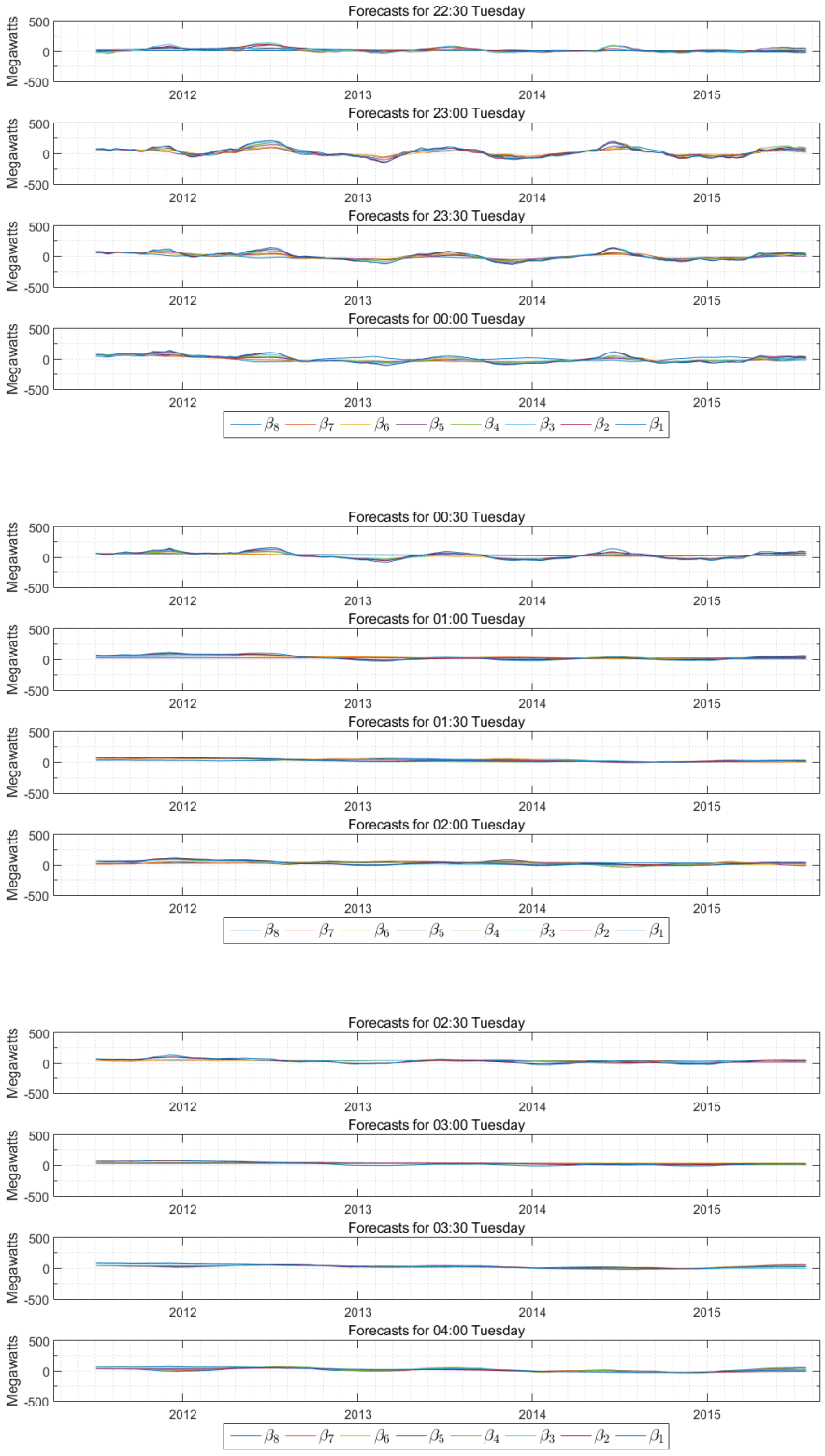


Figure D.9: Wednesday electricity trading day, 04:30 to 10:00 — estimates of bias in the demand forecasts.

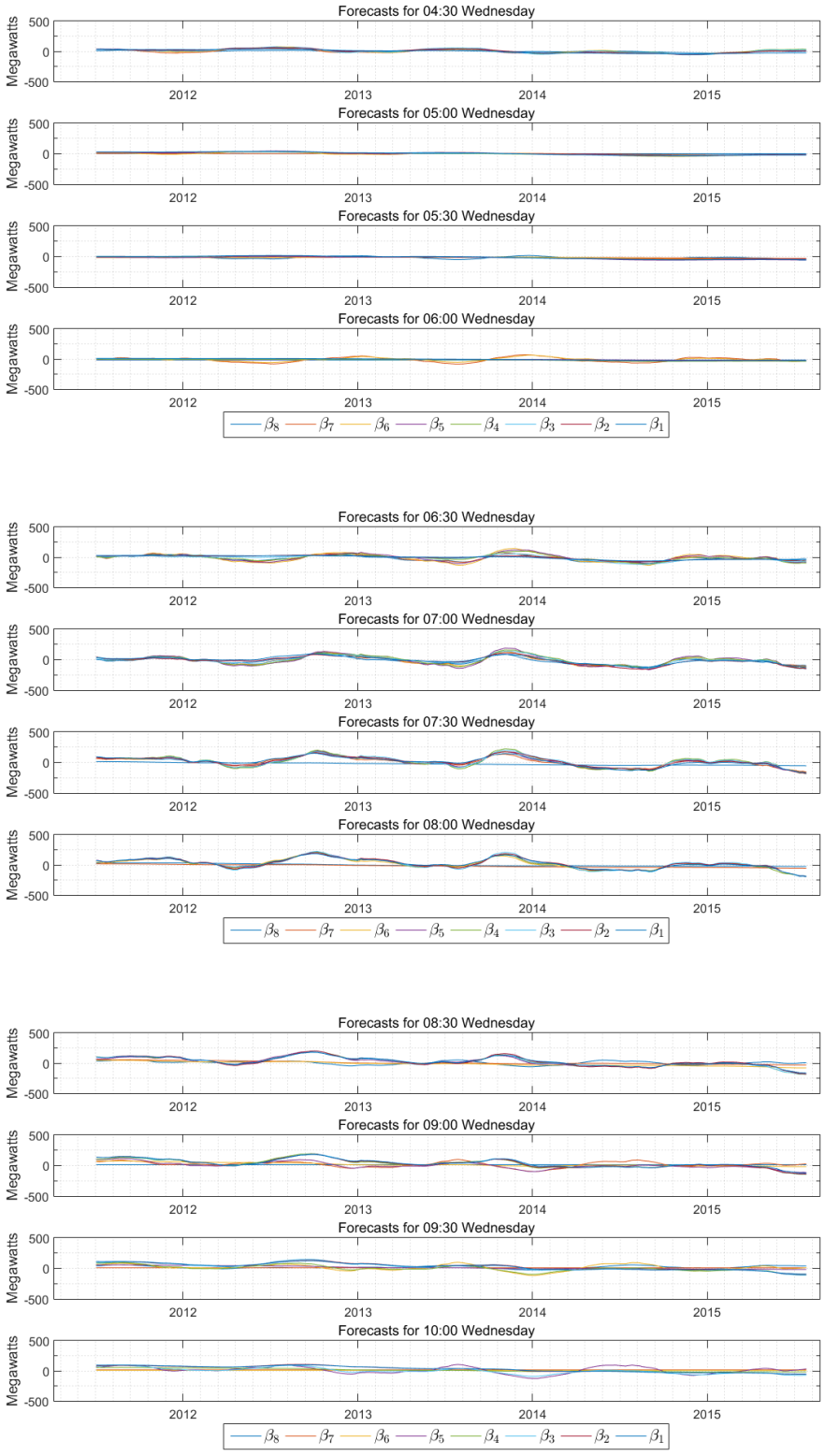


Figure D.10: Wednesday electricity trading day, 10:30 to 16:00 — estimates of bias in the demand forecasts.

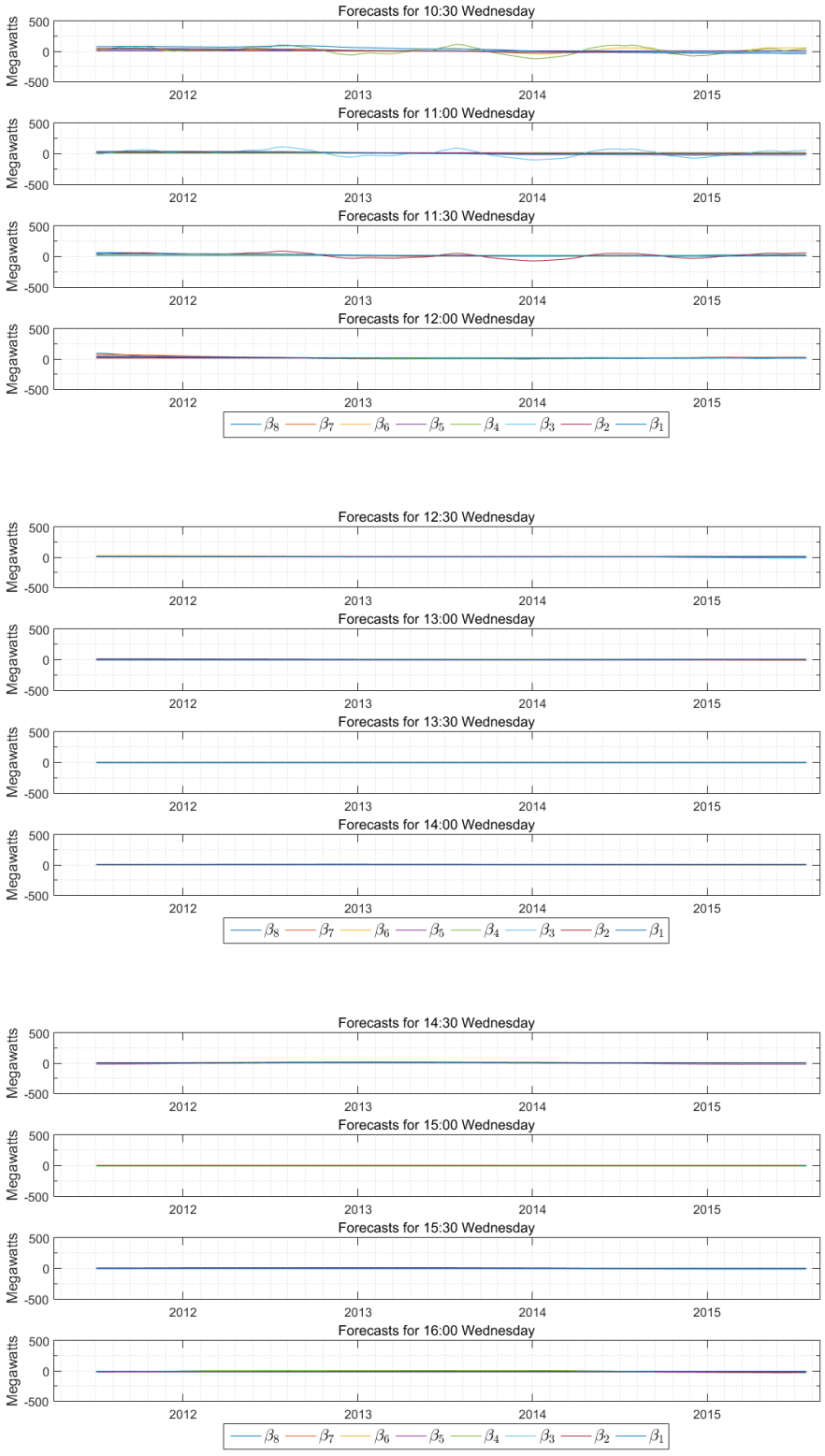


Figure D.11: Wednesday electricity trading day, 16:30 to 22:00 — estimates of bias in the demand forecasts.

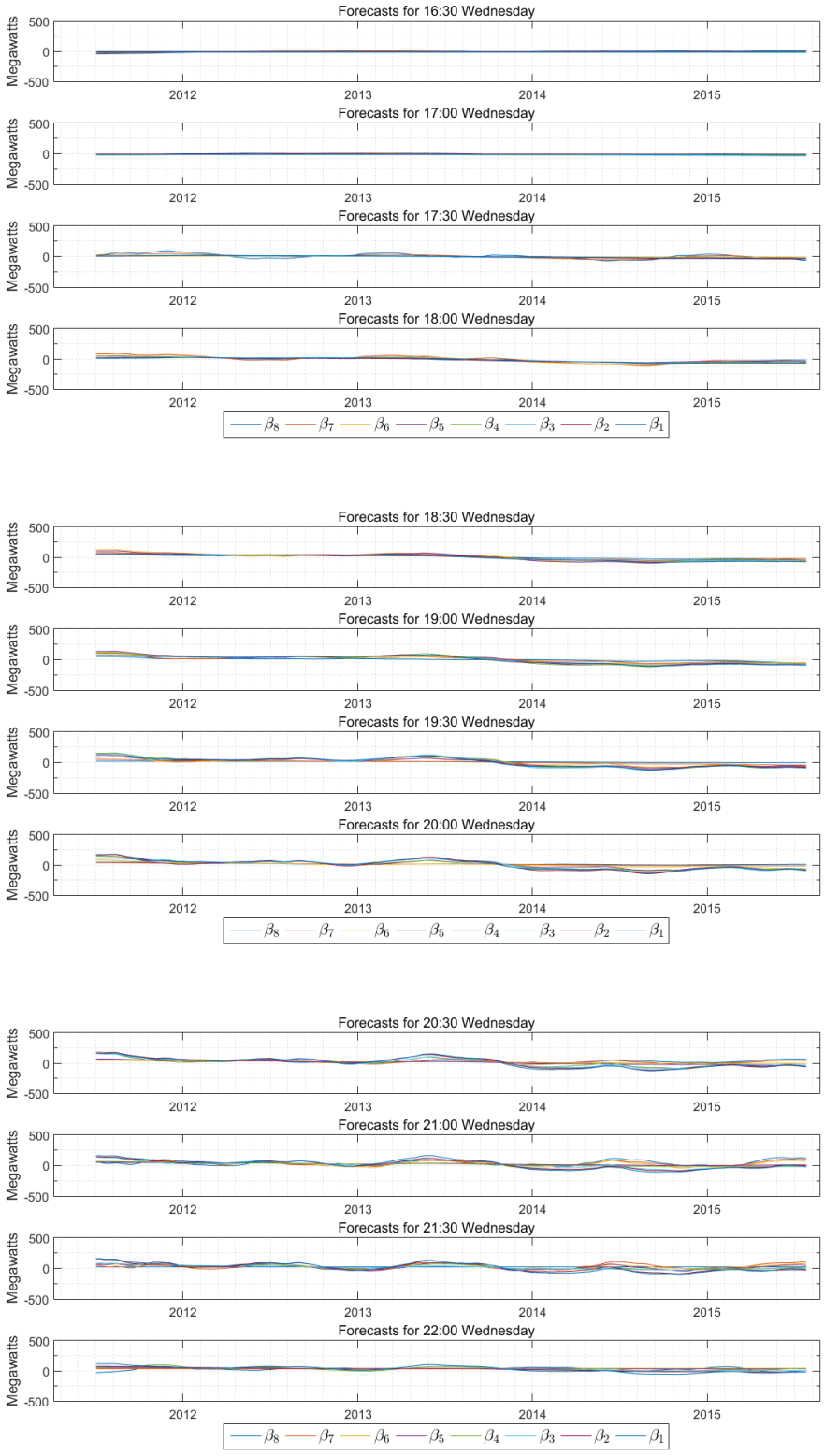


Figure D.12: Wednesday electricity trading day, 22:30 to 04:00 — estimates of bias in the demand forecasts.

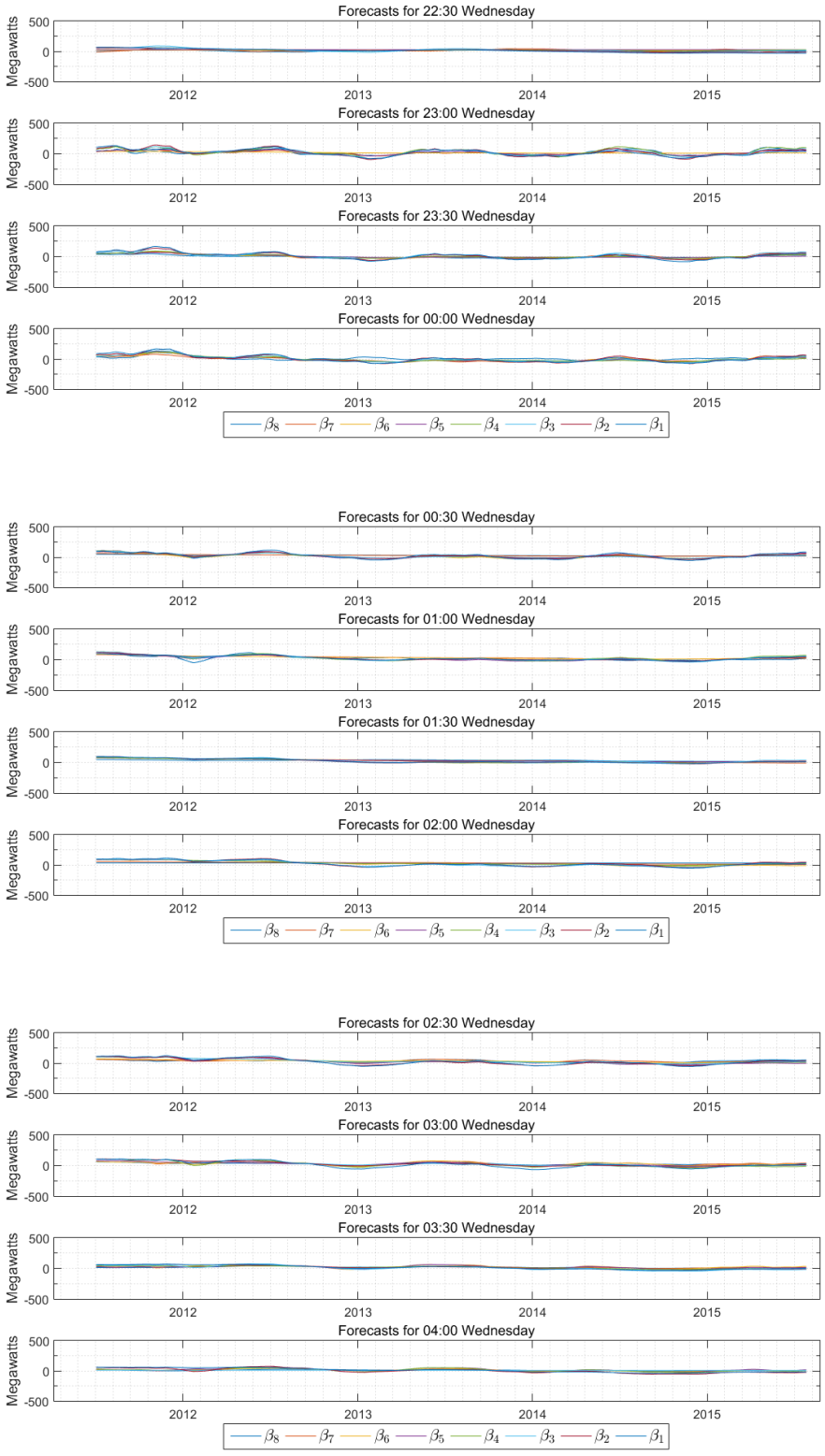


Figure D.13: Thursday electricity trading day, 04:30 to 10:00 — estimates of bias in the demand forecasts.

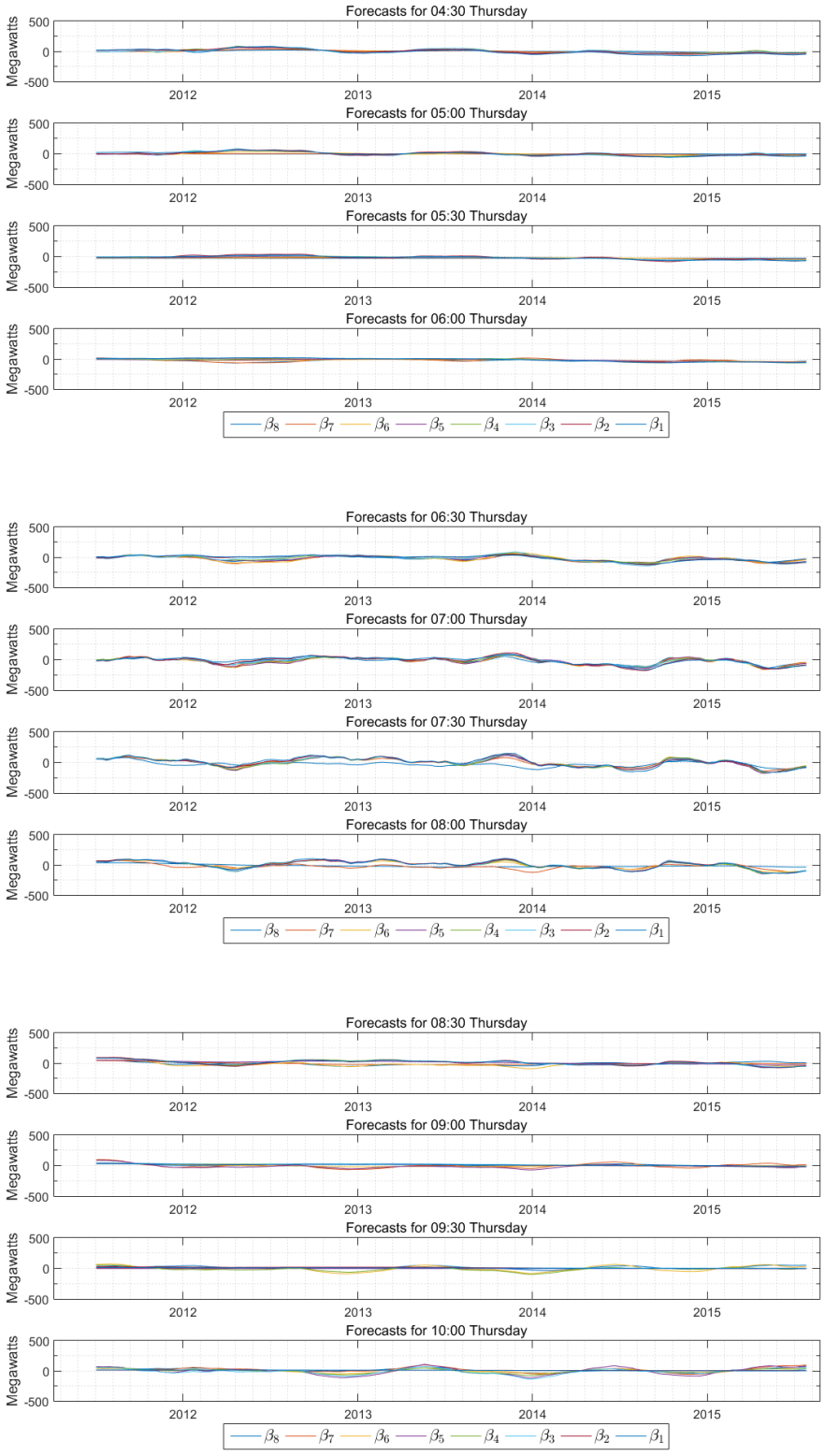


Figure D.14: Thursday electricity trading day, 10:30 to 16:00 — estimates of bias in the demand forecasts.

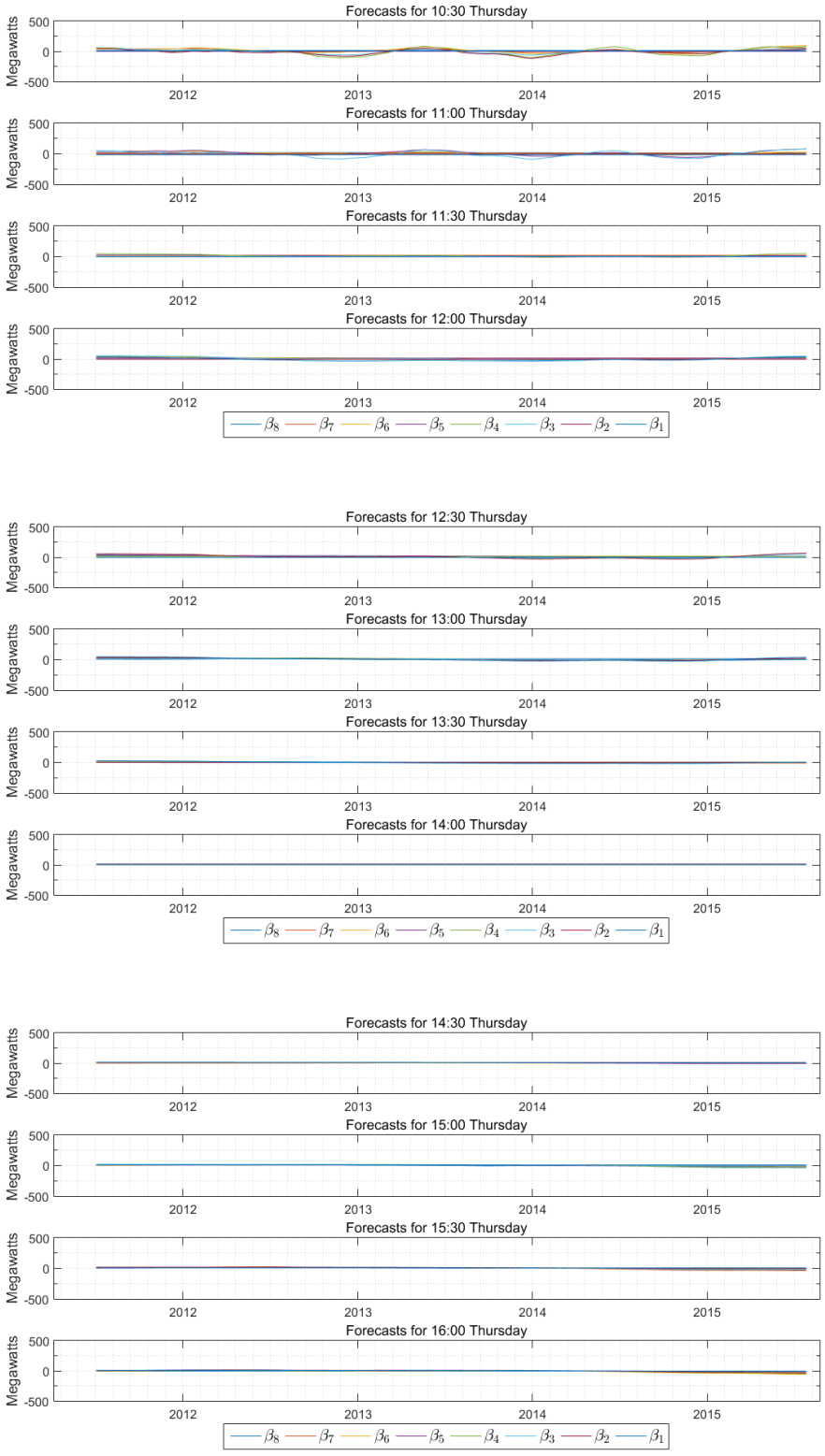


Figure D.15: Thursday electricity trading day, 16:30 to 22:00 — estimates of bias in the demand forecasts.

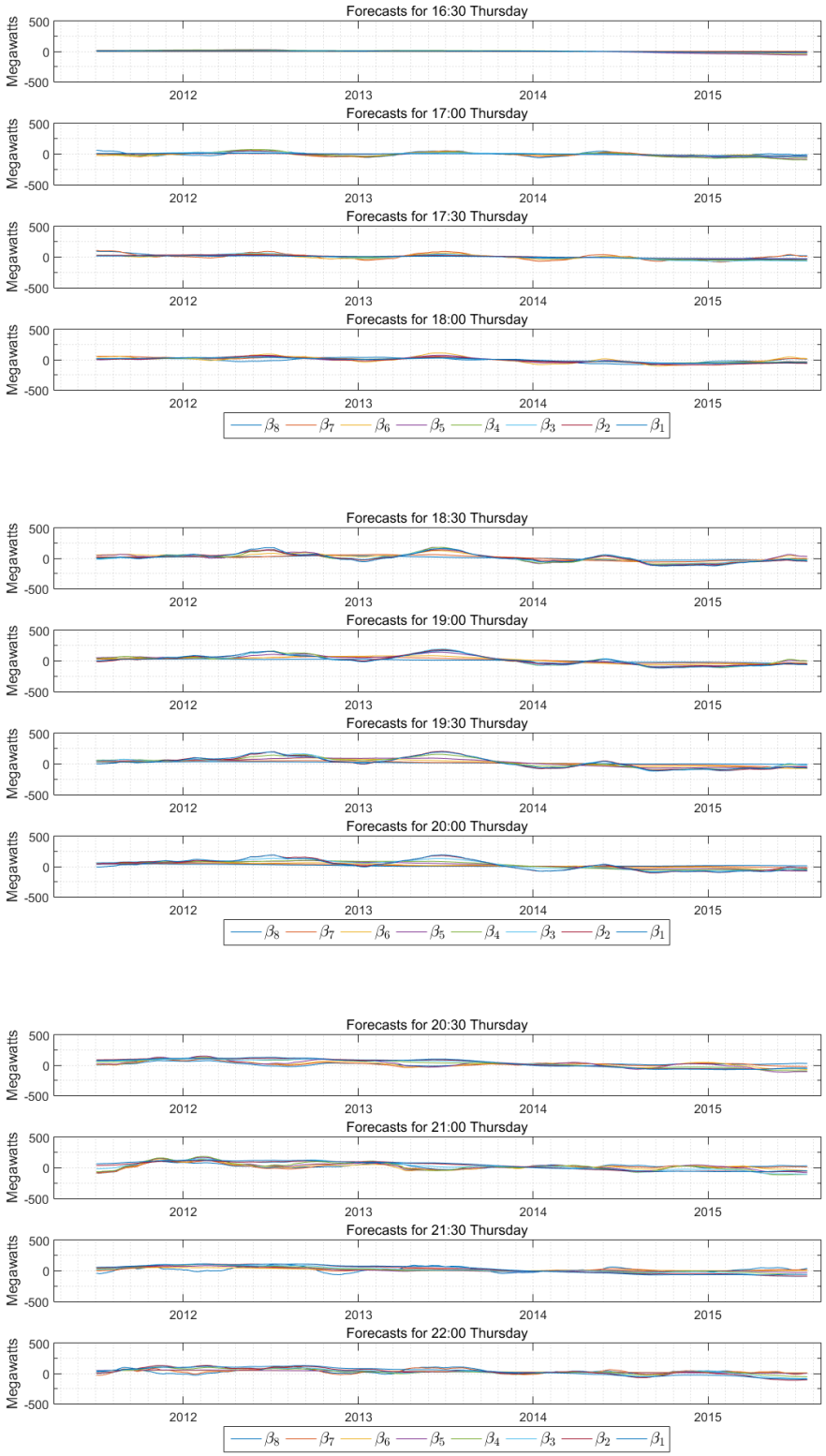


Figure D.16: Thursday electricity trading day, 22:30 to 04:00 — estimates of bias in the demand forecasts.

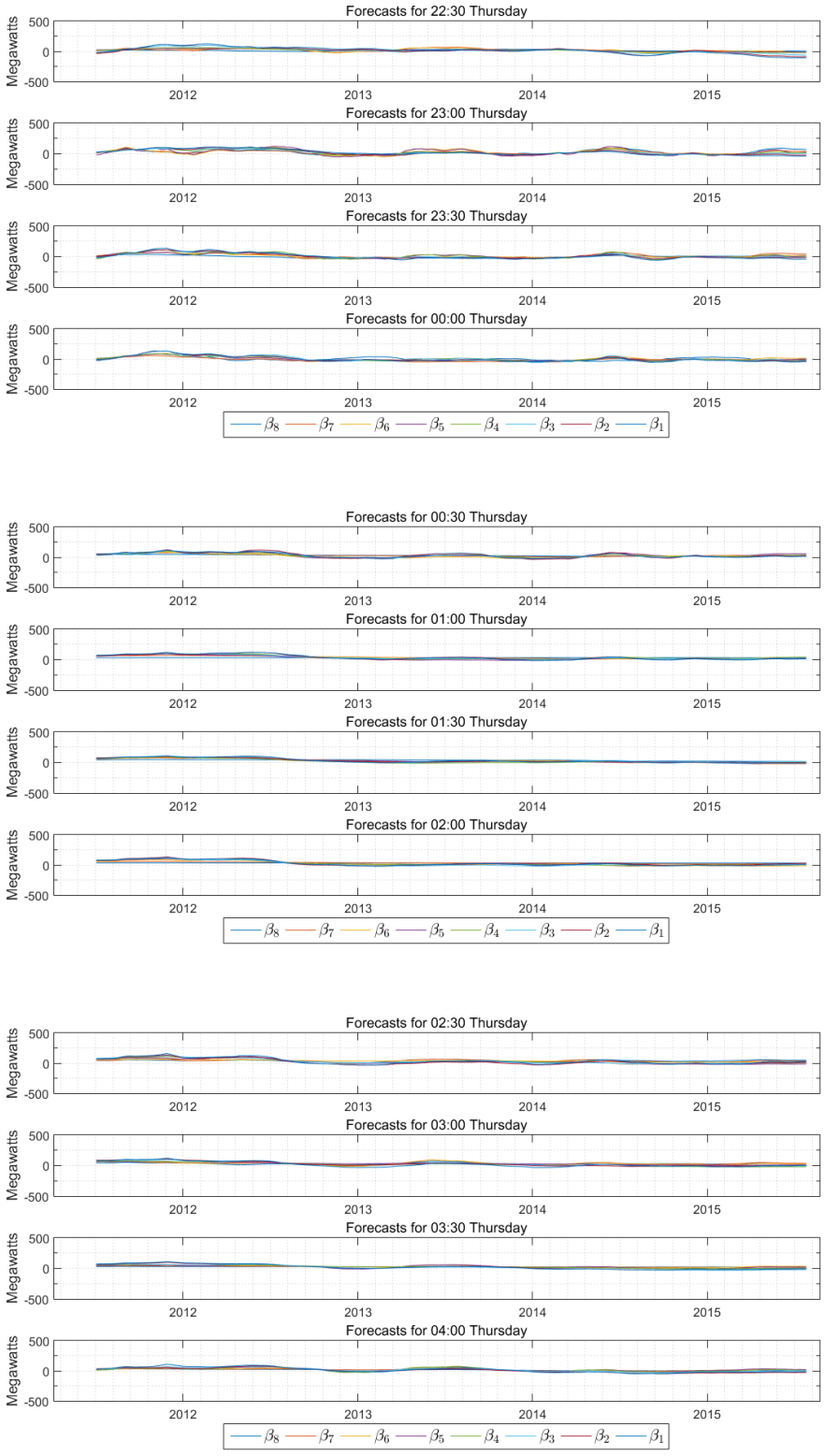


Figure D.17: Friday electricity trading day, 04:30 to 10:00 — estimates of bias in the demand forecasts.

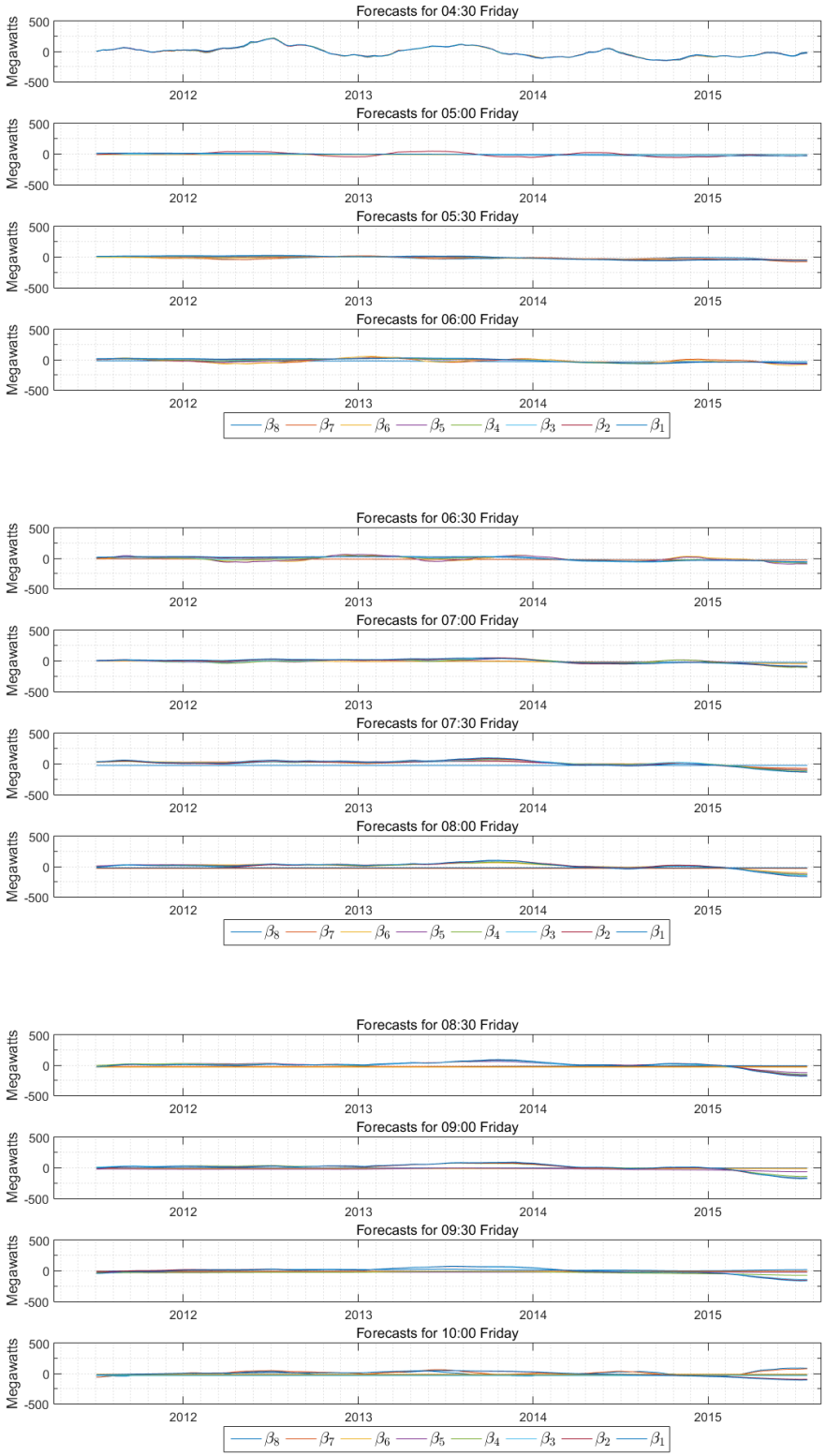


Figure D.18: Friday electricity trading day, 10:30 to 16:00 — estimates of bias in the demand forecasts.

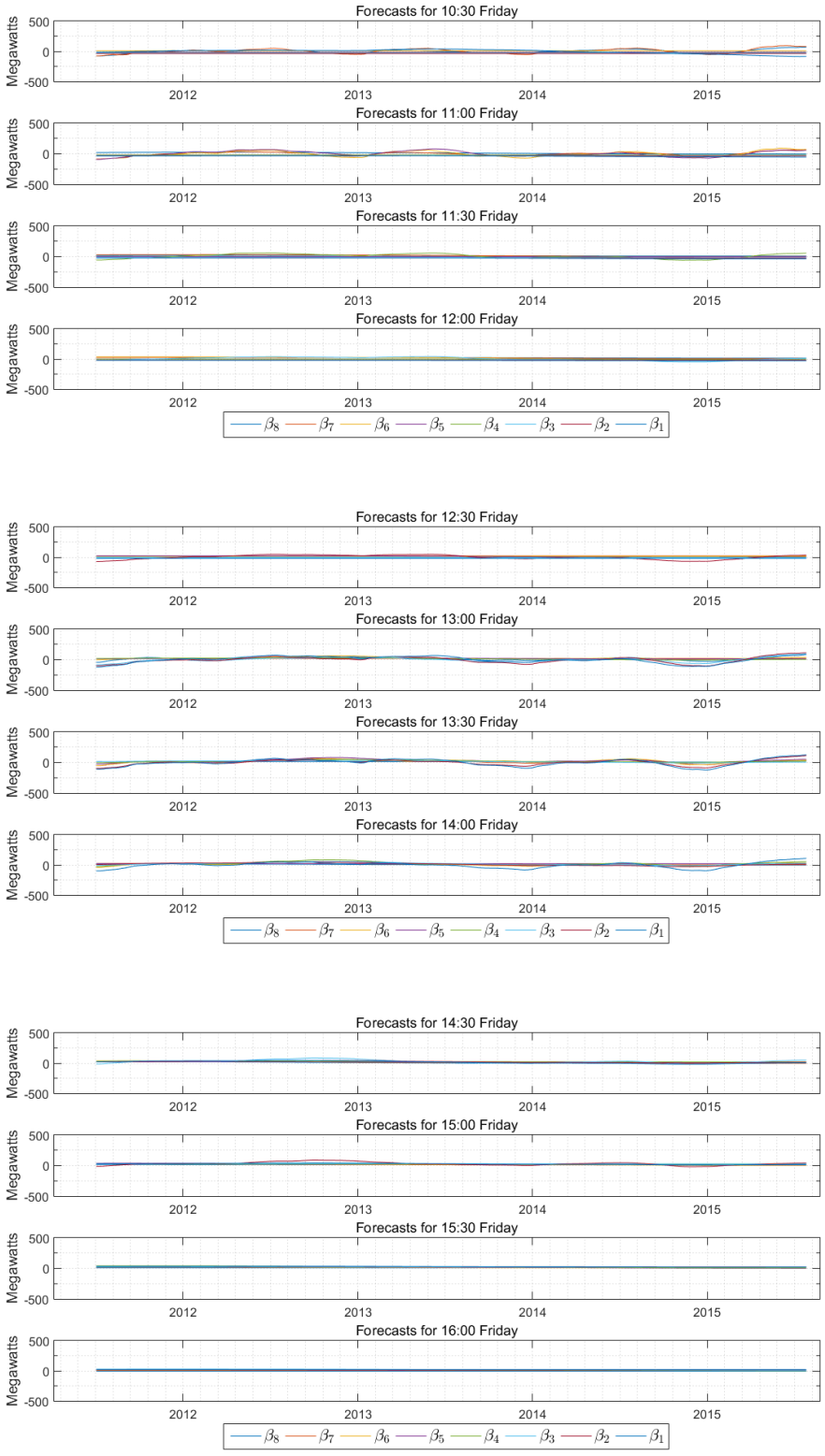


Figure D.19: Friday electricity trading day, 16:30 to 22:00 — estimates of bias in the demand forecasts.

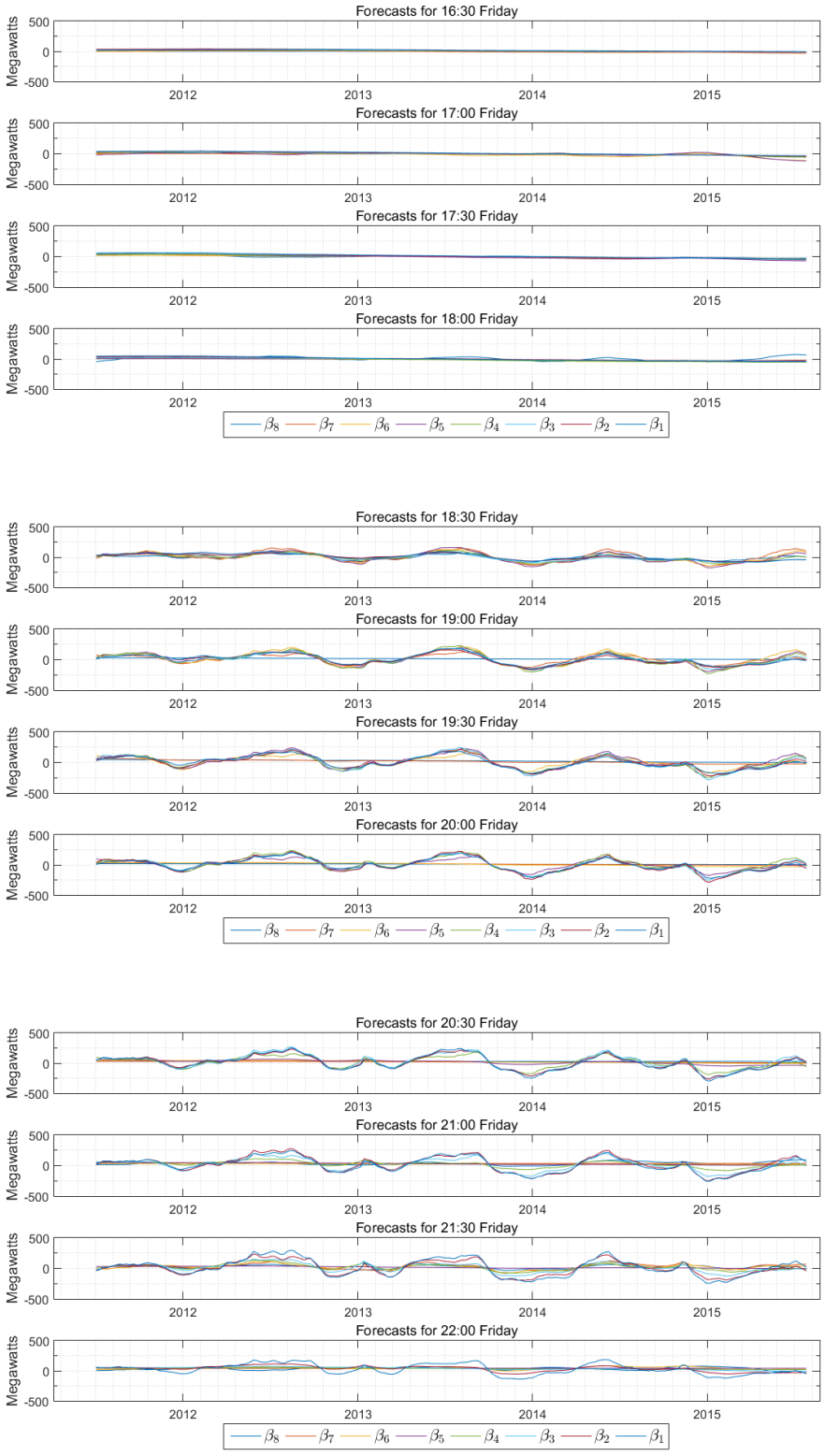


Figure D.20: Friday electricity trading day, 22:30 to 04:00 — estimates of bias in the demand forecasts.

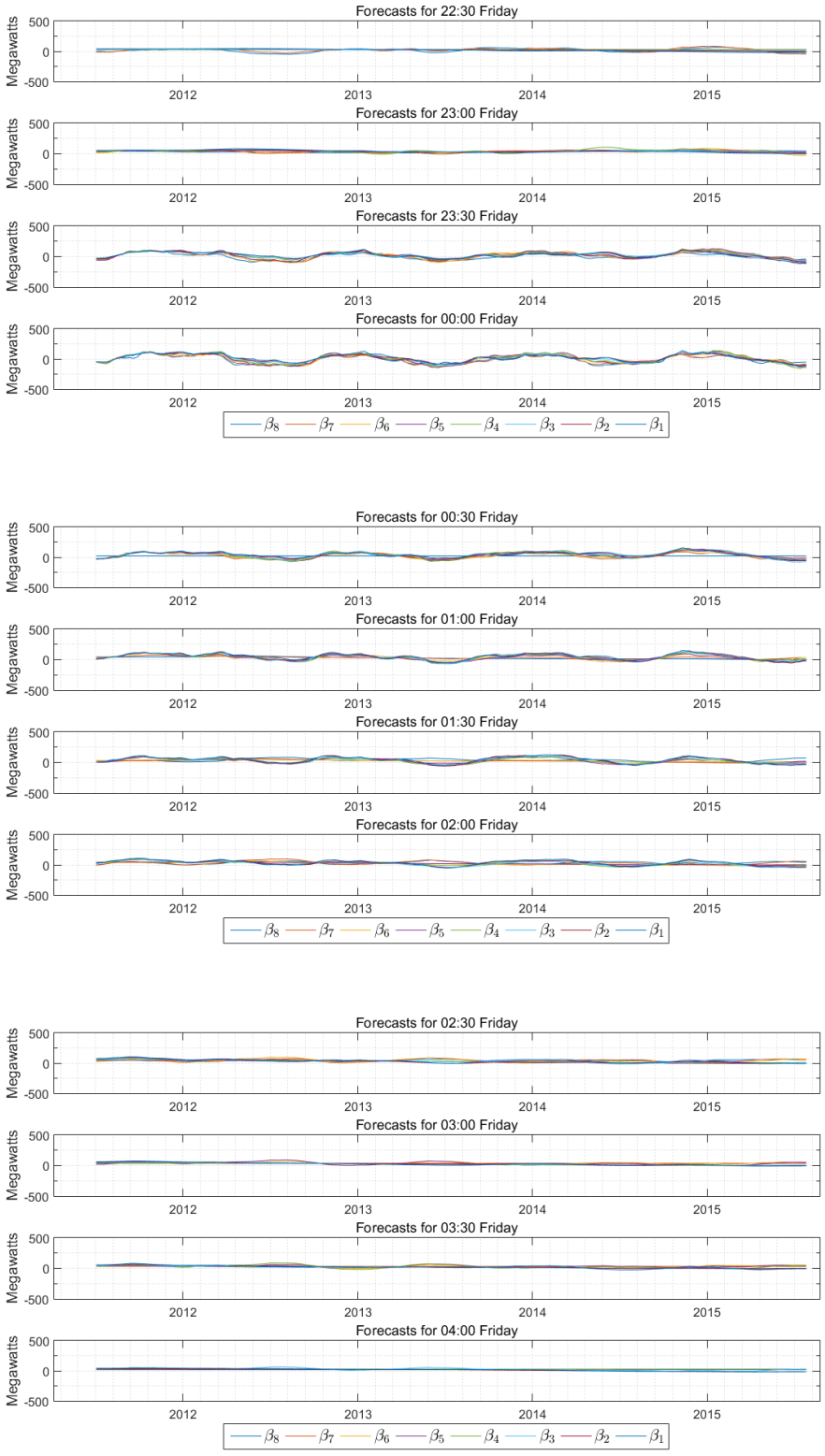


Figure D.21: Saturday electricity trading day, 04:30 to 10:00 — estimates of bias in the demand forecasts.

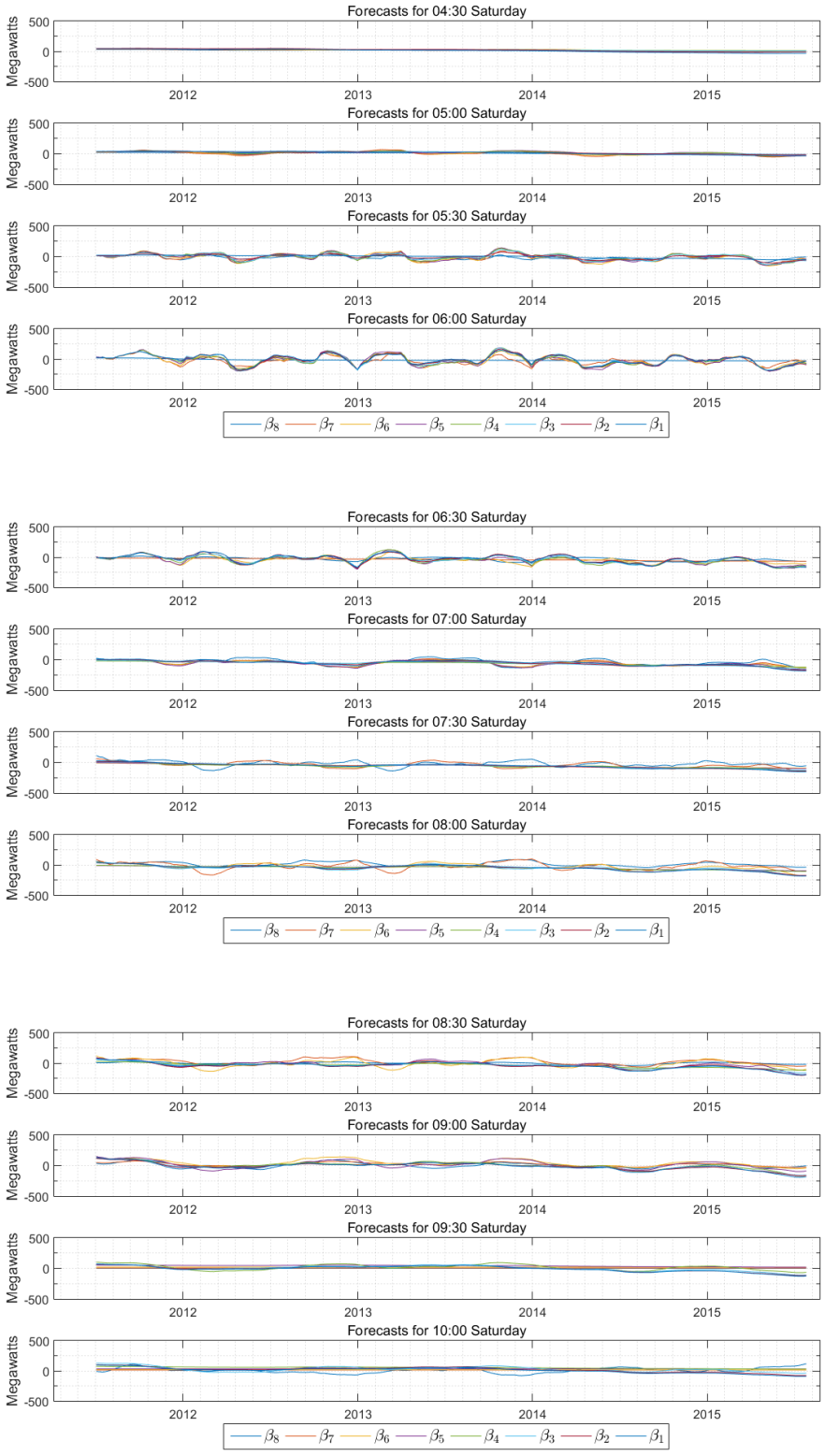


Figure D.22: Saturday electricity trading day, 10:30 to 16:00 — estimates of bias in the demand forecasts.

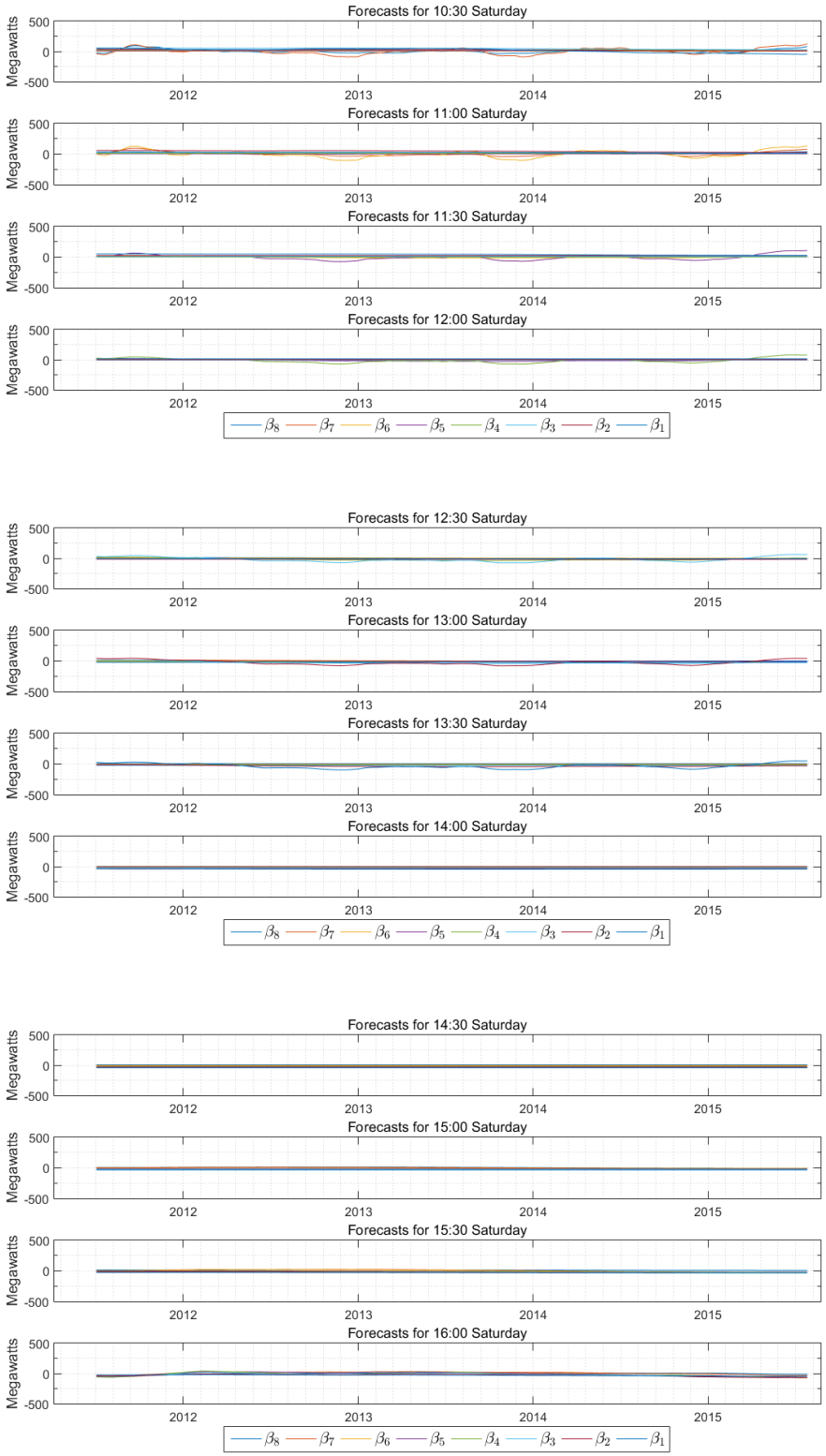


Figure D.23: Saturday electricity trading day, 16:30 to 22:00 — estimates of bias in the demand forecasts.

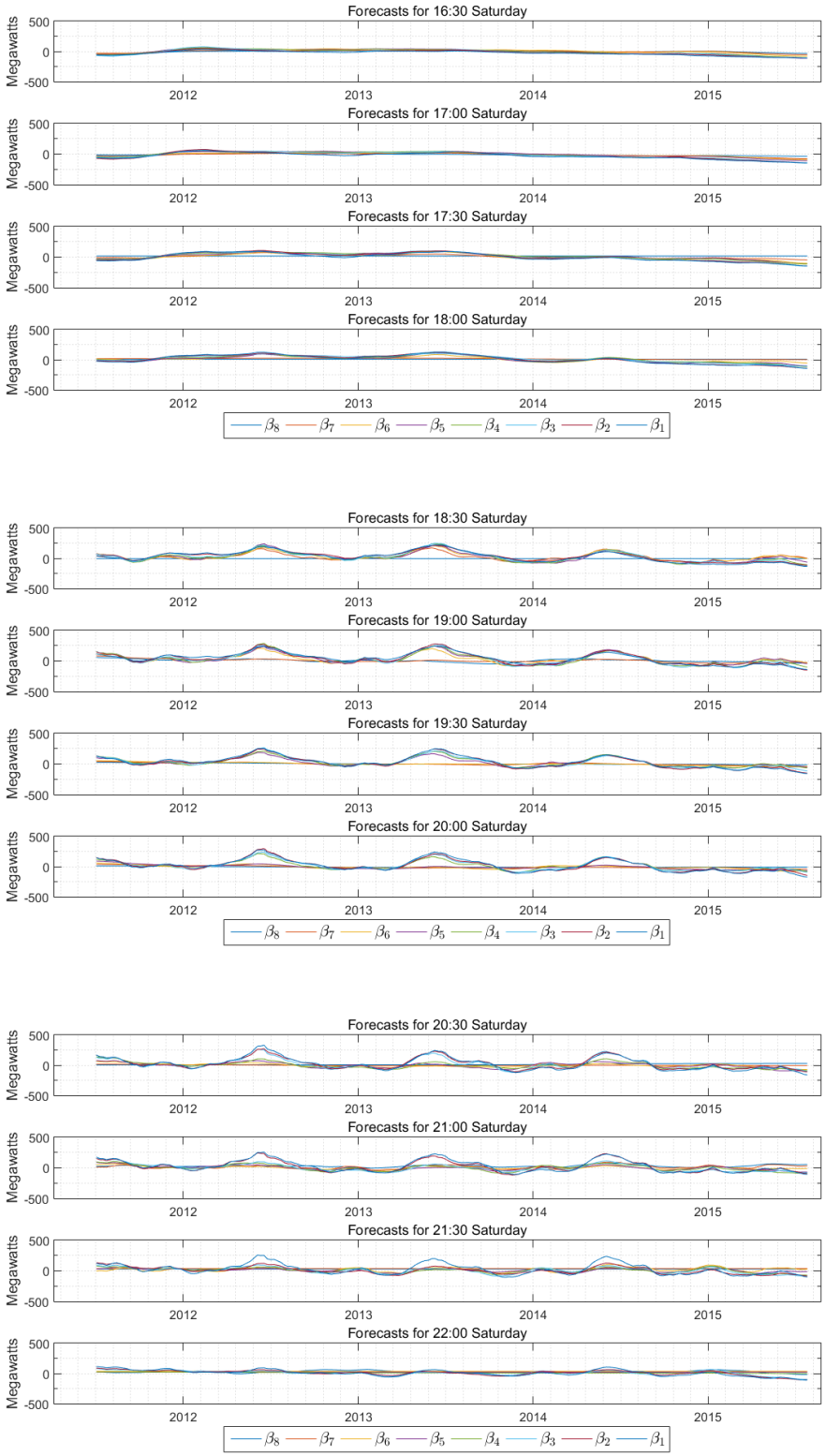


Figure D.24: Saturday electricity trading day, 22:30 to 04:00 — estimates of bias in the demand forecasts.

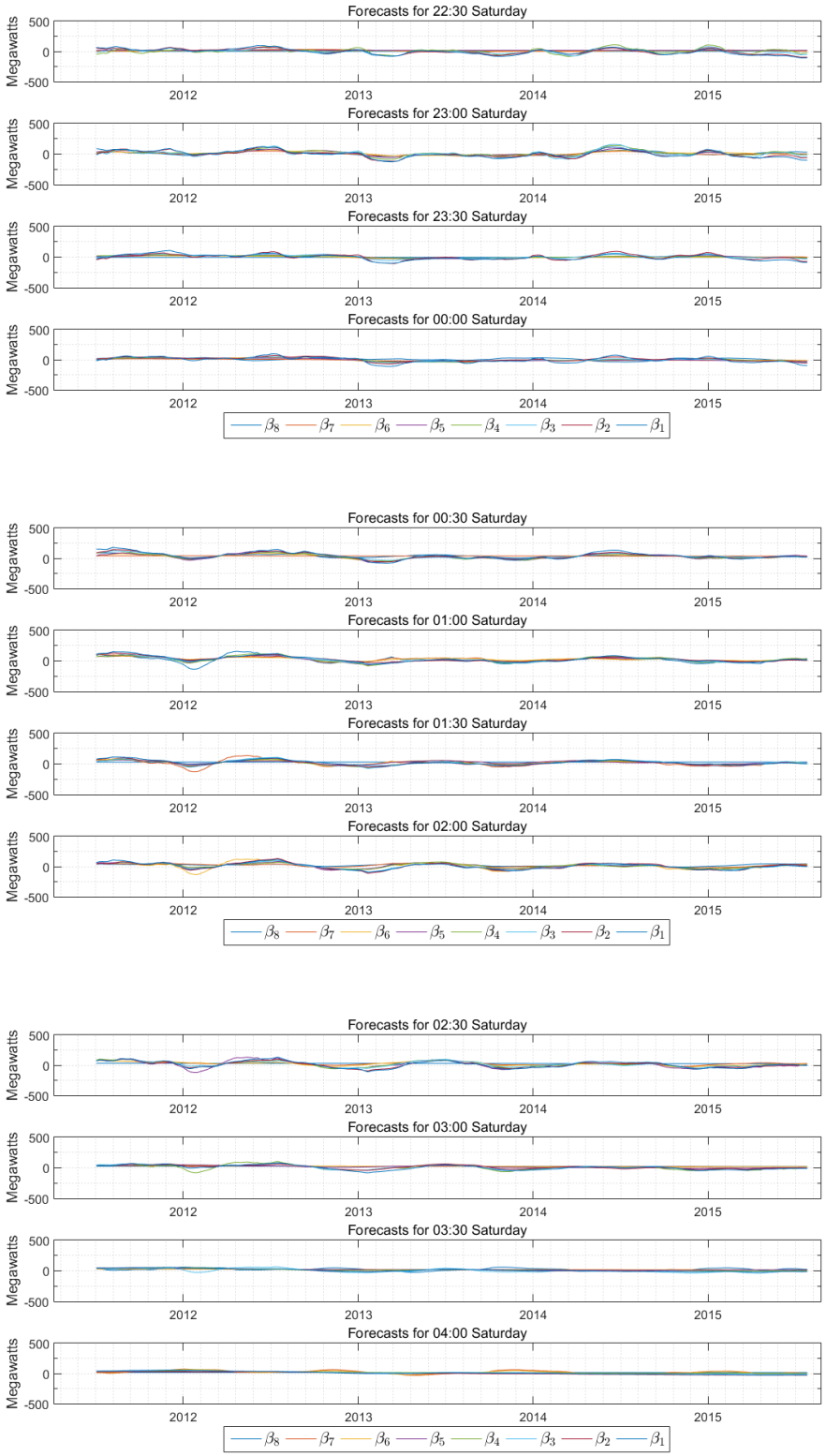


Figure D.25: Sunday electricity trading day, 04:30 to 10:00 — estimates of bias in the demand forecasts.

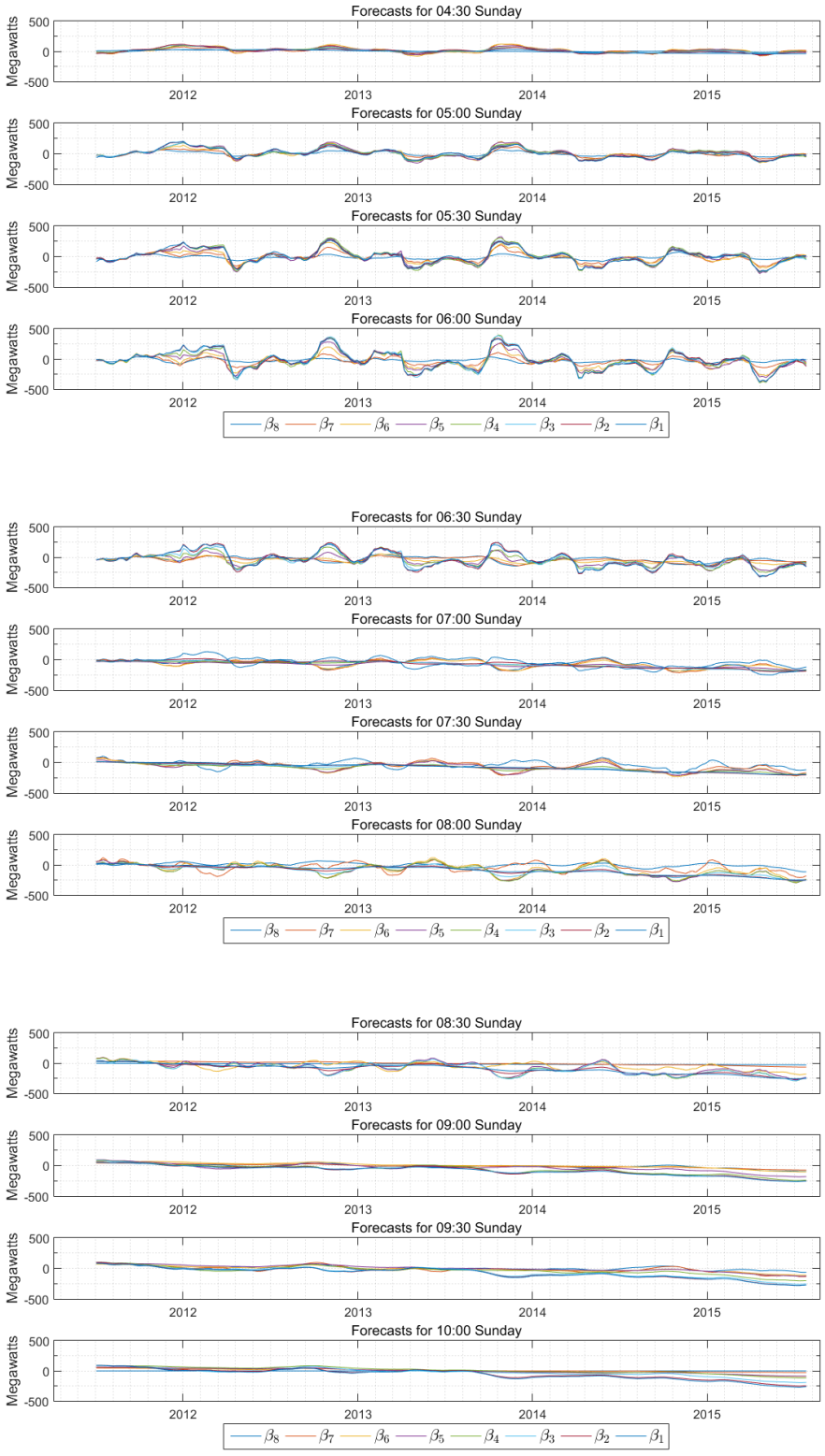


Figure D.26: Sunday electricity trading day, 10:30 to 16:00 — estimates of bias in the demand forecasts.

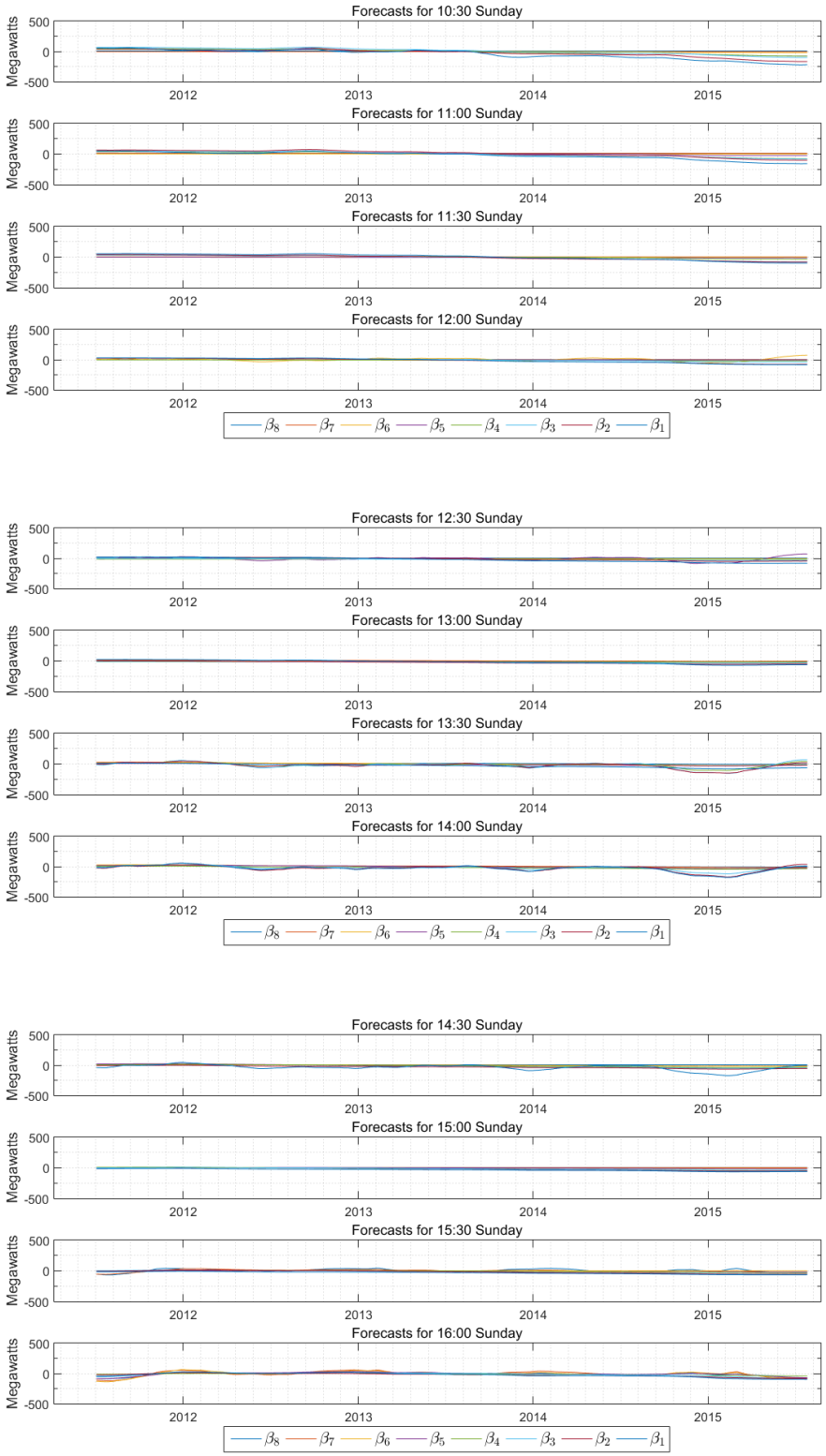


Figure D.27: Sunday electricity trading day, 16:30 to 22:00 — estimates of bias in the demand forecasts.

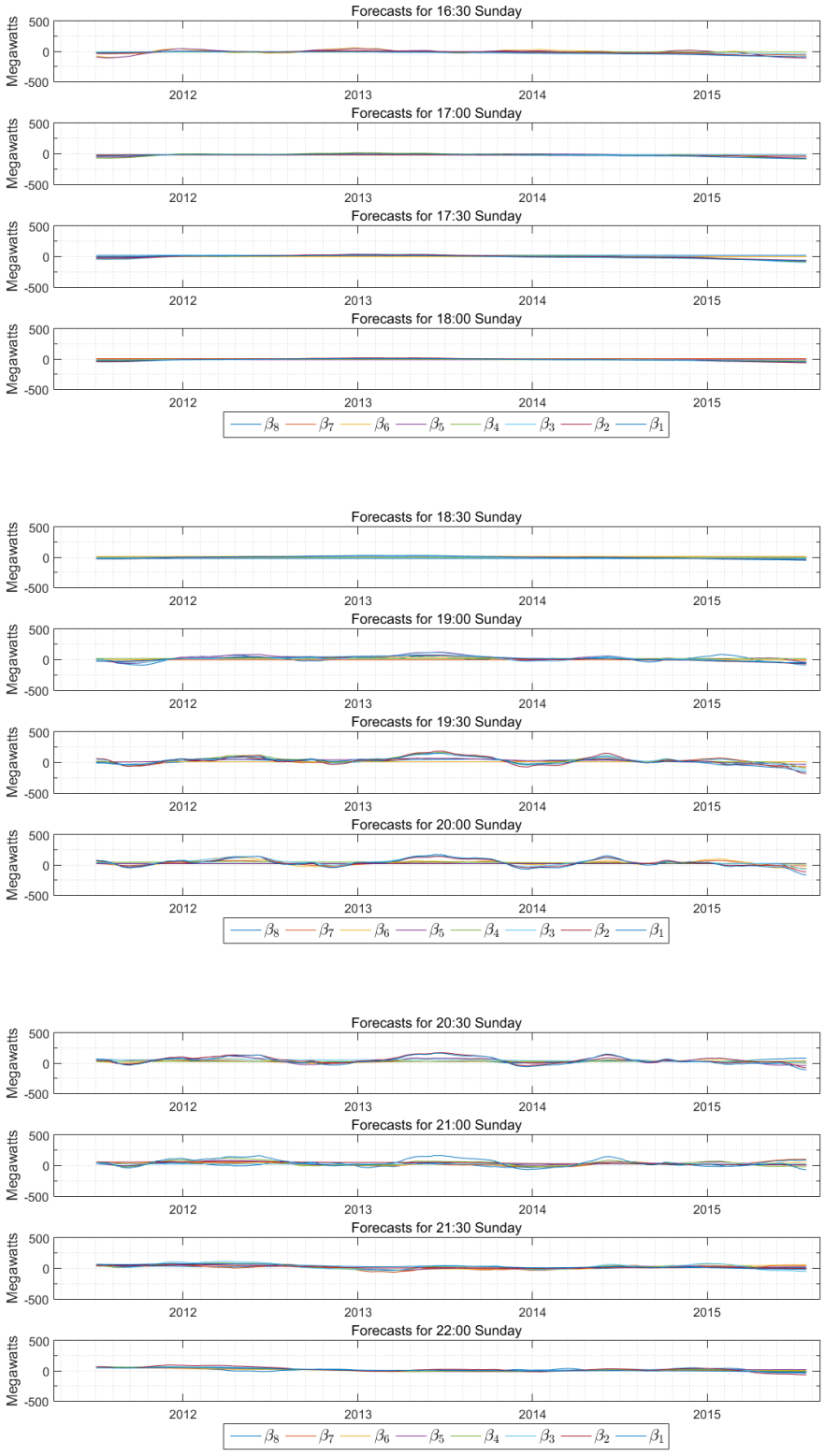
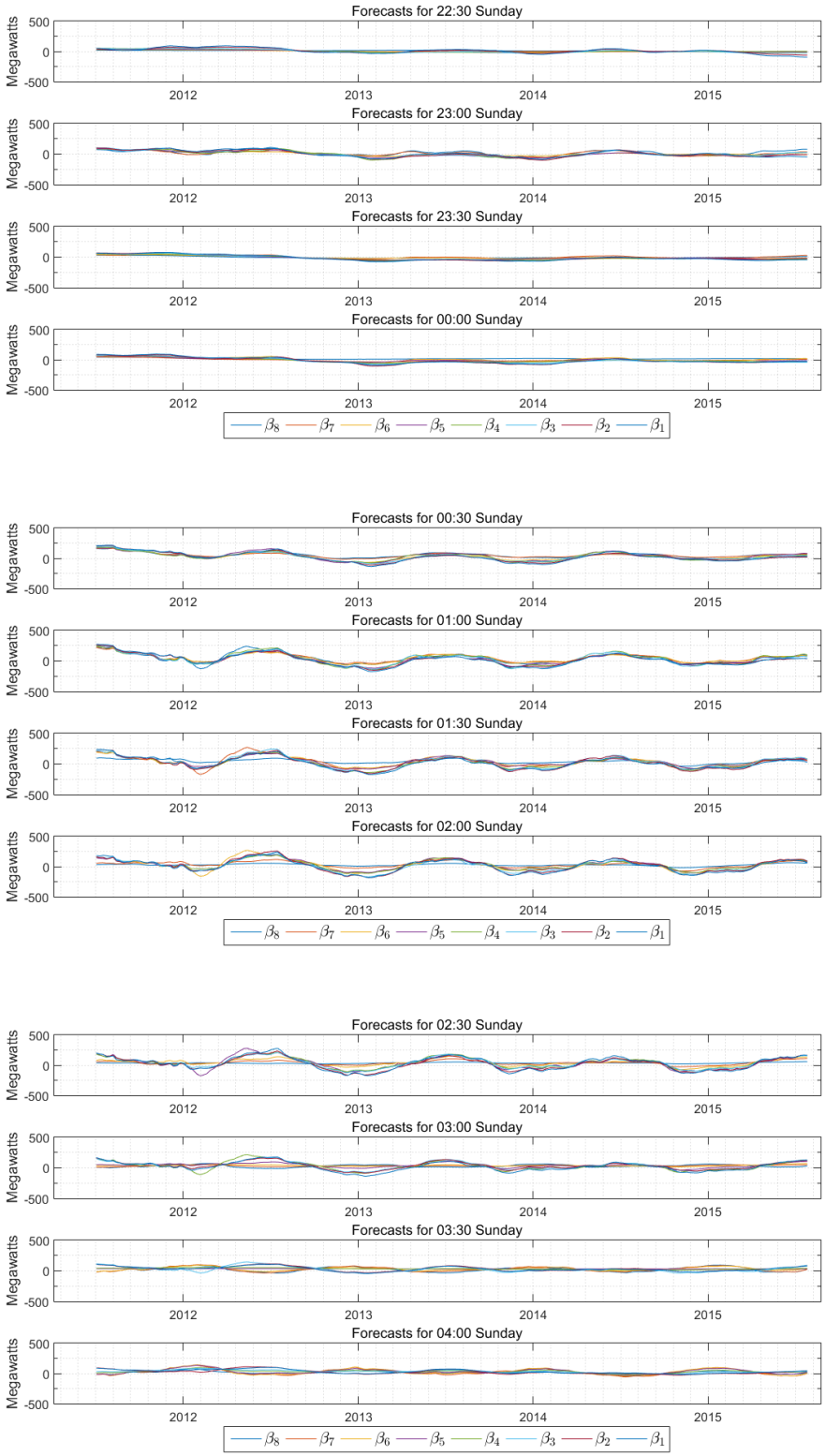


Figure D.28: Sunday electricity trading day, 22:30 to 04:00 — estimates of bias in the demand forecasts.



BIBLIOGRAPHY

AEMC. 2013. Potential generator market power in the NEM: Final rule determination. Technical report, 26 April 2013, Sydney.

AEMC. 2015a. Five minute settlement working group: Working Paper No. 1 — materiality of the problem and responsiveness of generation and load. Technical report, 12 October 2015, Sydney.

AEMC. 2015b. National electricity amendment (bidding in good faith) rule 2015: Final rule determination. Technical report, 10 December 2015, Sydney.

AEMO. 2014. Power system operating procedure: Load forecasting. Technical report, 11 June 2014, Sydney.

Aruoba SB. 2008. Data revisions are not well behaved. *Journal of Money, Credit and Banking* **40**: 319–340.

Baumeister C, Guérin P, Kilian L. 2015. Do high-frequency financial data help forecast oil prices? The MIDAS touch at work. *International Journal of Forecasting* **31**: 238–252.

Black JD, Henson WL. 2014. Hierarchical load hindcasting using reanalysis weather. *IEEE Transactions on Smart Grid* **5**: 447–455.

Clements MP. 1997. Evaluating the rationality of fixed-event forecasts. *Journal of Forecasting* **16**: 225–239.

Clements MP. 2005. *Evaluating econometric forecasts of economic and financial variables*. Palgrave Macmillan.

Clements MP, Galvão AB. 2008. Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business and Economic Statistics* **26**: 546–554.

Clements MP, Galvão AB. 2009. Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics* **24**: 1187–1206.

Clements MP, Taylor N. 2001. Robust evaluation of fixed-event forecast rationality. *Journal of Forecasting* **20**: 285–295.

Croushore D. 2011. Frontiers of real-time data analysis. *Journal of Economic Literature* : 72–100.

Croushore D. 2012. Comment. *Journal of Business and Economic Statistics* **30**: 17–20.

Davies A, Lahiri K. 1995. A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics* **68**: 205–227.

Davies A, Lahiri K. 1999. Re-examining the rational expectations hypothesis using panel data on multi-period forecasts. In *Analysis of panels and limited dependent variable models*. Cambridge University Press Cambridge, 226–254.

Davies A, Lahiri K, Sheng X. 2011. Analyzing three-dimensional panel data of forecasts. *The Oxford Handbook of Economic Forecasting* : 473–495.

Dell M, Jones BF, Olken BA. 2014. What do we learn from the weather? The new climate–economy literature. *Journal of Economic Literature* **52**: 740–798.

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.

Dreger C, Kholodilin KA. 2013. Forecasting private consumption by consumer surveys. *Journal of Forecasting* **32**: 10–18.

Duarte C, Rodrigues PM, Rua A. 2017. A mixed frequency approach to the forecasting of private consumption with ATM/POS data. *International Journal of Forecasting* **33**: 61–75.

Durbin J, Koopman SJ. 2012. *Time series analysis by state space methods*. Oxford University Press.

Fan S, Hyndman RJ. 2012. Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems* **27**: 134–141.

Fok D, Franses P, Paap R. 2007. Seasonality and non-linear price effects in scanner-data-based market-response models. *Journal of Econometrics* : 231–251.

Ghysels E, Ozkan N. 2015. Real-time forecasting of the US Federal Government budget: A simple mixed frequency data regression approach. *International Journal of Forecasting* **31**: 1009–1020.

Ghysels E, Santa-Clara P, Valkanov R. 2004. The MIDAS Touch: Mixed data sampling regression models. Technical report, Anderson Graduate School of Management, UCLA.

Ghysels E, Sinko A, Valkanov R. 2007. MIDAS regressions: Further results and new directions. *Econometric Reviews* **26**: 53–90.

Giacomini R, White H. 2006. Tests of conditional predictive ability. *Econometrica* **74**: 1545–1578.

Gourieroux C, Holly A, Monfort A. 1982. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica: journal of the Econometric Society* : 63–80.

Holden K, Peel DA. 1990. On testing for unbiasedness and efficiency of forecasts. *The Manchester School* **58**: 120–127.

Hong T, Wang P, White L. 2015. Weather station selection for electric load forecasting. *International Journal of Forecasting* **31**: 286–295.

Hsiang SM. 2016. Climate econometrics. Technical report, National Bureau of Economic Research.

Hurn AS, Silvennoinen A, Teräsvirta T. 2016. A smooth transition logit model of the effects of deregulation in the electricity market. *Journal of Applied Econometrics* **31**: 707–733.

Isiklar G, Lahiri K. 2007. How far ahead can we forecast? Evidence from cross-country surveys. *International Journal of Forecasting* **23**: 167–187.

Jacobs JP, Van Norden S. 2011. Modeling data revisions: Measurement error and dynamics of "true" values. *Journal of Econometrics* **161**: 101–109.

Jeong J, Maddala GS. 1991. Measurement errors and tests for rationality. *Journal of Business and Economic Statistics* **9**: 431–439.

Lahiri K. 2012. Comment. *Journal of Business and Economic Statistics* **30**: 20–25.

Lee CC, Chiu YB. 2011. Electricity demand elasticities and temperature: Evidence from panel smooth transition regression with instrumental variable approach. *Energy Economics* **33**: 896–902.

Lovell MC. 1986. Tests of the rational expectations hypothesis. *The American Economic Review* **76**: 110–124.

Mankiw NG, Runkle DE, Shapiro MD. 1984. Are preliminary announcements of the money stock rational forecasts? *Journal of Monetary Economics* **14**: 15–27.

Mankiw NG, Shapiro MD. 1986. News or noise: An analysis of GNP revisions. *Survey of Current Business* **66**: 20–25.

McCloskey DN. 1998. *The rhetoric of economics*. Univ of Wisconsin Press.

Miller SM. 1991. Forecasting federal budget deficits: How reliable are US Congressional budget office projections? *Applied Economics* **23**: 1789–1799.

Mills ES. 1957. The theory of inventory decisions. *Econometrica: Journal of the Econometric Society* **25**: 222–238.

Mincer JA, Zarnowitz V. 1969. The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. NBER, 3–46.

Monteforte L, Moretti G. 2013. Real-time forecasts of inflation: The role of financial variables. *Journal of Forecasting* **32**: 51–61.

Murphy AH. 1993. What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting* **8**: 281–293.

Muth JF. 1961. Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society* **29**: 315–335.

Muth JF. 1985. Properties of some short-run business forecasts. *Eastern Economic Journal* **11**: 200–210.

Nordhaus WD. 1987. Forecasting efficiency: Concepts and applications. *The Review of Economics and Statistics* : 667–674.

Patton AJ, Timmermann A. 2012. Forecast rationality tests based on multi-horizon bounds. *Journal of Business and Economic Statistics* **30**: 1–17.

Ramanathan R, Engle R, Granger CW, Vahid-Araghi F, Brace C. 1997. Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting* **13**: 161–174.

Ritter M, Musshoff O, Odening M. 2011. Meteorological forecasts and the pricing of temperature futures. *The Journal of Derivatives* **19**: 45–60.

Rossi B, Sekhposyan T. 2015. Forecast rationality tests in the presence of instabilities, with applications to Federal Reserve and survey forecasts. *Journal of Applied Econometrics* .

Sent EM. 2002. How (not) to influence people: the contrary tale of John F. Muth. *History of Political Economy* **34**: 291–319.

Siklos PL. 2008. What can we learn from comprehensive data revisions for forecasting inflation? Some US evidence. *Forecasting in the Presence of Structural Breaks and Model Uncertainty* **3**: 271.

Smith P. 2016. Google's MIDAS touch: Predicting UK unemployment with internet search data. *Journal of Forecasting* **35**: 263–284.

Stekler HO. 2002. The rationality and efficiency of individuals' forecasts. In *A Companion to Economic Forecasting*. Wiley-Blackwell, 222–240.

Stern H. 2006. Combining human and computer generated forecasts using a knowledge based system. In *22nd Conference on Interactive Information and Processing Systems*, volume 27.

Stern H. 2007. Improving forecasts with mechanically combined predictions. *Bulletin of the American Meteorological Society* : 449–851.

Stern H. 2008. The accuracy of weather forecasts for Melbourne, Australia. *Meteorological Applications* **15**: 65–71.

Stern H, Davidson NE. 2015. Trends in the skill of weather prediction at lead times of 1–14 days. *Quarterly Journal of the Royal Meteorological Society* **141**: 2726–2736.

Taieb SB, Hyndman RJ. 2014. A gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting* **30**: 382–394.

Taylor JW. 2008. An evaluation of methods for very short-term load forecasting using minute-by-minute British data. *International Journal of Forecasting* **24**: 645–658.

Taylor JW, Buizza R. 2003. Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting* **19**: 57–70.

Thompson GD, Wilson PN. 1999. Market demands for bagged, refrigerated salads. *Journal of Agricultural and Resource Economics* : 463–481.

Weron R. 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* **30**: 1030–1081.

West KD. 2012. Comment. *Journal of Business and Economic Statistics* **30**: 34–35.

White H. 2001. *Asymptotic theory for econometricians*. Academic press.

Wilks DS. 2011. *Statistical methods in the atmospheric sciences*. Academic press.

Wolak FA. 1987. An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association* **82**: 782–793.

Wolak FA. 1989. Testing inequality constraints in linear econometric models. *Journal of Econometrics* **41**: 205–235.

Wooldridge JM, White H. 1988. Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* **4**: 210–230.

Zainudin WNRA, Becker R, Clements A. 2015. The Australian electricity market's pre-dispatch process: Some observations on its efficiency using ordered probit model. *AIP Conference Proceedings* **1691**.

Zarnowitz V. 1985. Rational expectations and macroeconomic forecasts. *Journal of Business and Economic Statistics* **3**: 293–311.

Zavala VM, Constantinescu EM, Krause T, Anitescu M. 2009. On-line economic optimization of energy systems using weather forecast information. *Journal of Process Control* **19**: 1725–1736.

Zhang ZY, Gong DY, Ma JJ. 2014. A study on the electric power load of Beijing and its relationships with meteorological factors during summer and winter. *Meteorological Applications* **21**: 141–148.